ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8: RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.

Gerard Pujadas Anguiano

ISBN: 978-84-694-2163-5
Dipòsit Legal: T.885-2011

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8: RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

UNIVERSITAT ROVIRA I VIRGILI.

DEPARTAMENT DE BIOQUÍMICA I BIOTECNOLOGIA.

Estudis computacionals sobre β-amilasa, una molècula enzimàtica tipus barril $(\beta/\alpha)_8$: relacions evolutives i transicions estructurals al centre catalític.

Memòria presentada per

GERARD PUJADAS ANGUIANO

per optar al grau de Doctor en Ciències Químiques
sota la direcció del Dr. Jaume Palau Albet.

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8:
RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

UNIVERSITAT ROVIRA I VIRGILI.

DEPARTAMENT DE BIOQUÍMICA I BIOTECNOLOGIA.

Estudis computacionals sobre β-amilasa, una molècula enzimàtica tipus barril $(\beta/\alpha)_8$: relacions evolutives i transicions estructurals al centre catalític.

Memòria presentada per

GERARD PUJADAS ANGUIANO

per optar al grau de Doctor en Ciències Químiques
sota la direcció del Dr. Jaume Palau Albet.

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8: RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

UNIVERSITAT
ROVIRA I VIRGILI

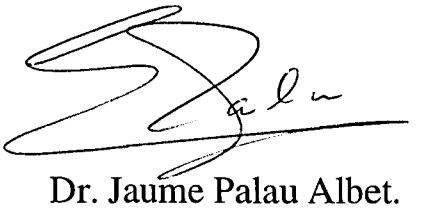
Prof. Jaume Palau
DEPARTAMENT DE BIOQUÍMICA I BIOTECNOLOGIA
Facultat de Química. Pça. Imperial Tàrraco, 1
43005-Tarragona. Catalonia. SPAIN
Tel. 34-77 55 95 65. Fax 34-77 55 95 97

Dr. Jaume Palau Albet, Catedràtic de Bioquímica i Biotecnologia Molecular del Departament de Bioquímica i Biotecnologia de la Universitat Rovira i Virgili,

FA CONSTAR,

que el present treball, amb el títol "**Estudis computacionals sobre β-amilasa, una molècula enzimàtica tipus barril (β/α)$_8$: relacions evolutives i transicions estructurals al centre catalític.**" que presenta en GERARD PUJADAS ANGUIANO, per optar al Grau de Doctor en Ciències Químiques per la Universitat Rovira i Virgili, ha estat realitzat sota la meva direcció, i que tots els resultats obtinguts són fruit dels experiments duts a terme per l'esmentat doctorand.

I perquè se'n prengui coneixement i tingui els efectes que correspongui, signa aquesta certificació.



Dr. Jaume Palau Albet.

Tarragona, març de 1998.

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8: RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

Taula de continguts.

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8: RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8: RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8: RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8:
RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

Introducció.

I. Característiques estructurals dels barrils $(\beta/\alpha)_8$.

I.1. El primer domini conegut amb plegament $(\beta/\alpha)_8$.

La primera proteïna cristal·litzada en que es va observar l'estructura de barril $(\beta/\alpha)_8$ va ésser l'enzim triosafosfatisomerasa (D-gliceraldehid 3-fosfat cetol isomerasa: EC 5.3.1.1) de múscul de pit de pollastre (*Gallus gallus*) (Banner et al., 1975). Aquest enzim, situat a la via d'Embden-Meyerhof-Parnas (glucòlisi), catalitza la reacció d'isomerització que permet la interconversió entre el fosfat de dihidroxiacetona (DHAP) i el D-gliceraldehid-3-fosfat (GAP).

L'anàlisi del mapa de densitat electrònica obtingut a una ressolució de 2.5 Å va demostrar que per a cada unitat asimètrica hi ha una molècula dimèrica. Aquest dímer està constituit per dues cadenes polipeptídiques de 247 residus cadascuna, químicament idèntiques. Els autors van descriure l'estructura de cada subunitat com a aproximadament esfèrica, sorprenentment regular i amb un diàmetre al voltant dels 35 Å. La cadena polipeptídica de cadascuna de les subunitats té un plegament intern en forma de barril on les dogues estan representades per les vuit tires d'una làmina β paral·lela. Aquestes tires, per efecte del cargolament que pateixen les làmines β *per se*, es troben inclinades respecte l'eix del barril. Les tires β paral·leles properes a nivell d'estructura 3-D són consecutives a nivell d'estructura primària. Intercalades amb les tires β, tant en la seqüència com en el plegament, hi ha els segments helicoidals que recobreixen al barril per la seva part externa.

Banner et al., (1975) van descriure que el plegament de cadascuna de les subunitats de l'enzim triosafosfatisomerasa tenia una estructura tipus $(\beta/\alpha)_8$. No obstant, cal assenyalar que els segments corresponents a les estructures secundàries no són idèntics i que el tros de seqüència que uneix dues tires β consecutives inclouen, en alguns casos, més d'una hèlice.

El fet que aquest enzim sigui conegut també per les sigles TIM (***Triose phosphate isomerase***) ha provocat que aquest plegament també sigui anomenat barril TIM (amb independència de quina sigui l'activitat enzimàtica que correspongui). Un altre sinònim que sovint s'empra és el de barril αβ.

ISBN:978-84-694-2163-5/DL:T-885-2011

I.2. Característiques generals del plegament.

Amb posterioritat als treballs efectuats per Banner et al. (1975), s'ha descobert que hi han d'altres proteïnes amb el mateix tipus de plegament que el descrit per l'enzim triosafosfatisomerasa de pollastre. Aquest conjunt de proteïnes, tot i ésser molt semblants en la seva aparença 3-D, no constitueixen un conjunt totalment homogeni. No obstant això, es poden definir una sèrie de característiques generals que corresponen al que seria el barril TIM canònic. Aquesta estructura representativa es caracteritza per la presència de vuit tires β paral·leles, en l'interior del domini, que es troben unides entre si mitjançant una sèrie de ponts d'hidrògen que interconnecten les veïnes. Per exemple, la tira β_1 està unida mitjançant ponts d'hidrògen a les tires β_2 i β_8, la β_2 a β_1 i β_3 i així succesivament. La làmina β resultant té una estructura d'hiperboloide producte de la proximitat en l'espai de les tires β_1 i β_8. El fragment de seqüència que separa dues tires β consecutives contindria una única hèlice α. Aquestes hèlices són majoritàriament amfipàtiques i es disposen espacialment al voltant de la làmina β_8 central recubrint-la. Les cares hidrofòbica i hidrofílica de les hèlices interaccionen amb la cara externa de la làmina β i amb el medi aquós respectivament.

Moltes proteïnes amb aquest plegament s'aparten lleugerament de l'estructura canònica $(\beta/\alpha)_8$. Com a exemples d'aquestes petites desviacions de la idealitat es poden destacar: a) la presència d'alguna tira β antiparal·lela respecte la resta [com en el cas de l'enolasa on les dues primeres tires β són antiparal·leles i el sentit de la primera hèlice és el contrari del de la resta (Lebioda et al., 1989)]; b) la manca d'alguna de les tires β que constitueixen el barril central [com en el cas de la cel·lobiohidolasa II que conté una barril format únicament per 7 tires β (Rouvinen et al., 1990)].

I.3. Paper dels diferents tipus d'elements d'estructura secundària en l'estabilitat del domini.

I.3.1. La làmina β paral·lela.

I.3.1.1. Característiques geomètriques.

Una de les característiques estructurals més importants dels barrils TIM és la presència d'una làmina β paral·lela que es troba situada al cor de l'estructura (vegeu secció I.2.). Normalment, aquesta làmina està constituïda per vuit tires β i es troba enrotllada sobre si mateixa.

ISBN:978-84-694-2163-5/DL:T.885-2011

L'estructura que en resulta es pot imaginar, *grosso modo*, com si fos una bastida en forma d'hiperboloide amb els eixos de les tires β situats a sobre de la seva superfície.

La primera descripció amb detall d'un barril format per una làmina β va ésser realitzada a principis de la decada dels 70 per Birktoft i Blow (1972). Posteriorment, diversos autors han intentat parametritzar i correlacionar les principals característiques geomètriques d'aquests tipus d'estructures (McLachlan, 1979; Richardson, 1981; Salemme i Weatherford, 1981a,b; Salemme, 1981; Chothia i Janin, 1982; Novotny et al., 1984; Lasters et al., 1988; Lesk et al., 1989; Chothia i Finkelstein, 1990; Lasters, 1990; Lasters et al., 1990; Wodak et al., 1990; Ke, 1992; Murzin et al., 1992; Murzin et al., 1994a,b). En general, tant les relacions com els paràmetres s'han obtingut a partir de barrils ideals, és a dir, amb forma de cilindre o d'hiperboloide de secció circular. La principal desviació respecte a aquesta "idealitat" es deguda a l'existència de barrils amb una secció elíptica (Lasters et al., 1988; Murzin et al., 1994a,b). És important assenyalar que tant els paràmetres com les equacions emprats a l'hora de definir aquesta geometria no són exclusius de les estructures amb plegament de barril TIM sinó que, en general, són vàlids per a qualsevol barril constituït per una làmina β tancada (amb independència del nombre de tires β que hi participin i de si la seva disposició és o no és paral·lela) (McLachlan, 1979; Murzin et al., 1994a,b). A continuació es presenten quins són els paràmetres necessaris a l'hora de descriure la geometria d'un barril constituït per una làmina β (entre parèntesi i al final de la definició hi figura el valor corresponent al paràmetre en un barril TIM ideal):

Paràmetres intrínsecs a la pròpia làmina β:

a: distància entre els C^{α} al llarg de les tires β (3.3 Å).

b: distància entre tires β veïnes (4.4 Å).

τ: valor promig del cargolament de la làmina β respecte a un eix perpendicular a la direcció de cada tira β (-26°).

θ: valor promig del cargolament de la làmina β respecte a un eix paral·lel a la direcció de cada tira β.

ε: valor promig de l'enrotllament de la làmina β al llarg de cada tira β.

μ: valor promig de l'enrotllament de la làmina β al llarg d'una línia perpendicular a les tires β.

Paràmetres característics d'una determinada estructura en forma de barril:

n: nombre de tires β que constitueixen el barril (8).

α: valor promig de l'angle que forma la direcció de cadascuna de les tires β amb l'eix del barril (-36°).

R: radi promig del barril (14 Å).

S: "desplaçament" de la làmina. Si les tires β fossin paral·leles a l'eix del barril (α = 0°), llavors qualsevol línia vertical a aquest eix que sortís d'un determinat residu de la primera tira β i continués a través de tota la làmina, tornaria exactament al punt de partida. En el cas dels barrils $(\beta/\alpha)_8$, la inclinació de les tires β fa que quan aquesta línia torna a la primera tira, ho faci a vuit residus de la posició inicial. Per tant, en el cas particular dels barrils TIM, **S** agafa el valor de 8.

Una gran part de tots aquests paràmetres ja van ser proposats el 1979 per McLachlan. Aquest autor va ser el primer en tractar de parametritzar les característiques geomètriques d'un barril i relacionar-les mitjançant les següents equacions:

$$S = \frac{n * b * \tan(\alpha)}{a} \qquad [1]$$

$$R = \frac{b}{2 * \sin(\pi/n) * \cos(\alpha)} \qquad [2]$$

Chothia (1988) va proposar una nova formulació d'aquestes dues equacions, que és vàlida únicament pels barrils $(\beta/\alpha)_8$ (s'obtenen de substituir **a** i **b** a [1] i [2] pels valors que els hi corresponen en els barrils TIM ideals):

$$\tan\alpha = \frac{0.75 * S}{n} \qquad [1]$$

$$R = \frac{2.2}{\sin(\pi/n) * \cos(\alpha)} \qquad [2]$$

Cal remarcar que les equacions de McLachlan van ser desenvolupades inicialment amb l'intenció d'aplicar-les a barrils constituïts per làmines β antiparal·leles. No obstant això, també són vàlides a l'hora d'estudiar la geometria de la làmina β paral·lela dels barrils $(\beta/\alpha)_8$ (Chothia 1988; Lesk et al., 1989).

Amb posterioritat, Lasters (1990) va voler aprofundir sobre quin és l'efecte de l'hiperboloide sobre el cargolament de la làmina β i quines son les restriccions conformacionals que li imposa la seva geometria. En aquest estudi, Lasters va introduir dos nous conceptes, el de "cargolament de la bastida" o T_s i el de "cargolament de la tira β" o $T_{w\beta}$ que, posteriorment, serien desenvolupats més ampliament per Murzin et al. (1994a) en descriure el cargolament i l'enrotllament de la làmina β (tot i que emprant diferents paràmetres (vegeu el següent paràgraf). El mateix any, Wodak et al. (1990) van demostrar que la làmina β dels barrils TIM no introdueix cap restricció extra sobre la cadena polipeptídica, és a dir, els parells d'angles díedres ϕ i ψ no presenten valors especials respecte als d'altres làmines β.

Més recentment, Murzin et al. (1994a) han desenvolupat unes equacions que extenen la parametrització a tota una sèrie de característiques estructurals com ara: a) la disposició de les seves tires β; b) la forma de l'hiperboloid; c) el nivell de cargolament i enrotllament de la làmina β; d) el valor òptim de cargolament necessari per a produir una làmina β tancada; i e) la disposició dels residus en l'interior del barril. En formular noves equacions, aquests autors van afegir nous paràmetres estructurals (θ, ε, τ i μ; vegeu definicions en un paràgraf anterior) als que prèviament ja havien estat definits per McLachlan (1979). Els resultats obtinguts per Murzin et al. (1994a) demostren que totes les propietats estructurals dels barrils venen determinades pels valors de **n** i de **S** i que, per tant, aquests dos valors poden ser emprats a l'hora de classificar els barrils. En concret, α i **R** depenen de **n** i de **S** a partir de les equacions proposades per McLachlan (equacions [1] i [2]). D'altra banda, θ, ε, τ i μ depenen de **n** i de **S** mitjançant les equacions:

$$n * \theta + S * \varepsilon = 2 * \pi * \sin(\alpha) \qquad [3]$$

$$S*(a/b)*\theta+n*\mu=2*\pi*\cos(\alpha) \qquad [4]$$

$$\tau=(a/b)*\theta \qquad [5]$$

A més, Murzin et al. (1994a) van suggerir que, de totes les combinacions possibles de valors per a **n** i **S**, únicament n'hi ha 10 que produeixen barrils amb un bon empaquetament de cadenes laterals al seu interior. Això indicaria que les proteïnes amb un barril format per una làmina β presenten un nombre molt limitat de possibles arquitectures. Lògicament, entre aquestes 10 combinacions hi ha la corresponent als barrils TIM (**n** i **S** = 8). Aquests autors també van estudiar com varien les equacions [1]-[5] si es considera que els barrils tenen una secció elíptica (i per tant, més propera al que està descrit per la majoria de barrils reals).

I.3.1.2. L'empaquetament dels residus a l'interior del barril.

L'empaquetament de cadenes laterals a l'interior de la làmina β ha estat descrit com a un dels factors més importants —sinó el que més— en el manteniment de l'estructura dels barrils TIM (Lesk et al., 1989; Houbrechts et al., 1995) i la seva major o menor termoestabilitat (Hwang et al., 1997). Normalment, els residus que interaccionen en el centre d'un barril són hidrofòbics però també es pot donar que n'hi hagi una certa quantitat d'hidrofílics (Helin et al., 1995).

La primera descripció sobre com interaccionen les cadenes laterals dels residus que es troben a la part central del barril va ser realitzada per en Lesk et al. (1989). Aquests autors van estudiar quina era la disposició dels residus dins del barril de tres enzims: la triosafosfatisomerasa (pollastre), el glicolatoxidasa (espinac) i la rubisco (*Rhodospirillum rubrum*). L'objectiu del seu estudi consistia en tractar d'explicar l'origen evolutiu dels barrils $(\beta/\alpha)_8$ des del punt de vista estructural (i no de seqüència que és com normalment es fa; vegeu la secció V.1. d'aquesta introducció). L'estudi va demostrar que en els tres casos hi ha 12 residus dins el cor de la làmina β i que la disposició de les seves cadenes laterals no és aleatòria. En totes tres s'observa com les seves cadenes laterals es reparteixen en tres plans que són —aproximadament— perpendiculars a l'eix del barril. Cadascun d'aquest plans o capes consta de 4 residus que pertanyen a

tires β alternades. A aquests plans se'ls ha qualificat com a "capes d'interacció" degut a que la majoria d'interaccions al centre del barril involucren residus que formen part d'una mateixa capa. Lesk et al. (1989) van trobar que tant en la triosafosfatisomerasa com en la rubisco, la capa central estava formada per residus de les tires β senars i les capes dels extrems per les tires β parelles. En el cas del glicolatoxidasa però, és a l'inrevés. En Lesk et al. van interpretar aquest doble tipus d'empaquetament des del punt de vista evolutiu (vegeu la secció V.1. d'aquesta introducció). Posteriorment, l'empaquetament a l'interior del barril ha estat estudiat tant de forma experimental (Tsuji et al., 1993) com teòrica (Murzin et al., 1994a,b). Els estudis realitzats per aquests darrers autors van demostrar que existeix una correlació entre els valors de **n** i **S** i la disposició dels residus a l'interior del barril (vegeu la secció I.3.1.1. d'aquesta introducció).

I.3.2. Les hèlices α.

Les hèlices α que es troben als barrils TIM són majoritàriament amfipàtiques i es disposen espacialment al voltant de la làmina β_8 central recubrint-la. Vtyurin i Panov (1995) han descrit que dels dos empaquetaments hidrofòbics en que participa la làmina β —un al seu interior i l'altre al seu exterior— el prioritari és l'extern (làmina β –hèlice α). Això seria conseqüència del major nombre de residus hidrofòbics en la interfase hèlices–làmina β (al voltant de 50) que no pas a l'interior del barril (al voltant de 12). Vtyurin i Panov (1995) han demostrat que els índex òptims d'empaquetament hidrofòbic —a banda i banda de la làmina— s'assoleixen quan el quocient entre **a** (distància promig entre els C^{α} alternats al llarg de les tires β) i **b** (distància promig entre tires β veïnes) és $\sqrt{2}$.

Nota: les definicions d'**a** i de **b** s'han bescanviat respecte a les que figuren a la referència original (Vtyurin i Panov, 1995) per a mantenir la coherència amb les emprades per McLachlan (1979), Chothia (1988) i Murzin et al. (1994a,b) (vegeu secció I.3.1.1. d'aquesta introducció).

I.3.3. Els llaços.

I.3.3.1. Els llaços de l'extrem N-terminal del barril.

Els llaços de l'extrem N- o amino terminal del barril són els que uneixen les hèlices α amb les tires β. El centre actiu dels enzims amb plegament de barril TIM es troba en l'extrem C-terminal del barril. Això vol dir que els llaços N-terminals no participen ni en l'activitat enzimàtica ni en la unió dels substrats o inhibidors a l'enzim. Si es comparen les seqüències d'un mateix enzim i diferent font es pot veure com el nivell de conservació de la seva seqüència és menor que el corresponent a la resta de l'estructura (Pujadas et al., 1996). Per tant, quin és el paper que juguen aquests llaços?. Poden tenir qualsevol tipus de seqüència sempre i quan no generi una estructura secundària?. Si no poden participar en la catàlisi perquè es troben lluny del centre actiu, participen del manteniment de l'estructura de l'enzim?. Les investigacions d'en Urfer i Kirschner (1992) han permés donar una resposta satisfactòria a aquestes preguntes. Urfer i Kirschner van utilitzar el gen d'un barril $(\beta/\alpha)_8$ —l'enzim monomèric fosforibosilantranilatisomerasa d'*E. coli*— per a construir i expressar una variant que contenia una duplicació interna —i en tandem— del cinquè motiu βα. La disposició relativa de les estructures secundàries en la variant era $\beta_1\alpha_1...\beta_5\alpha_5\beta'_5\alpha'_5...\beta_8\alpha_8$. Un cop la proteïna $(\beta/\alpha)_9$ va ser expressada i purificada, es va analitzar la seva activitat enzimàtica. Els resultats van demostrar que el mutant presentava activitat fosforibosilantranilatisomerasa i que, per tant, havia de mantenir el nucli $(\beta/\alpha)_8$ original. Això descartava la possibilitat de tenir un barril TIM on $\beta_1\alpha_1$ o $\beta_8\alpha_8$ fossin fora del nucli $(\beta/\alpha)_8$ del mutant. A més, tant $\beta_1\alpha_1$ com $\beta_8\alpha_8$ contenen residus que participen en la catàlisi i que, per tant , no poden allunyar-se de l'extrem C-terminal del barril. L'única manera doncs de conservar el nucli $(\beta/\alpha)_8$ original passa perquè un motiu αβ o βα del tandem $\beta_5\alpha_5\beta'_5\alpha'_5$ formi part d'un llaç. Això únicament pot passar si el motiu αβ o βα es col·loca en tres posicions diferents: a) entre α_4 i β'_5 (llaç a l'extrem N-terminal del barril format per $\beta_5\alpha_5$); b) entre α_5 i β_6 (llaç a l'extrem N-terminal del barril format per $\beta'_5\alpha'_5$); i finalment, c) entre β_5 i α'_5 (llaç a l'extrem C-terminal del barril format per $\alpha_5\beta'_5$). Mitjançant una sèrie d'experiències —independents unes de les altres— com ara

dicroisme circular, espectre d'emissió fluorescent, proteolisi limitada amb la proteasa V8, etc., Urfer i Kirschner (1992) van demostrar que de les tres possibilitats de plegament, només se'n produïa una que és la que té el motiu αβ entre β_5 i α'_5 ($\alpha_5\beta'_5$ situats en un llaç de l'extrem C-terminal del barril). La conclusió que en van extreure és que els llaços de l'extrem N-terminal del barril —i les zones properes a ells— són, en general, molt més importants per a l'estabilitat de l'estructura que no pas els de l'extrem C-terminal. Aquesta seria la raó per la qual el mutant prefereix tenir un motiu αβ prop del centre actiu —tot i que fa disminuir la k_{cat} 25 cops respecte a l'enzim natural— que no pas a la banda amino terminal del barril.

I.3.3.2. Els llaços de l'extrem C-terminal del barril.

Els llaços de l'extrem C- o carboxi terminal del barril són els que uneixen les tires β amb les hèlices α. Aquests llaços —a més del seu paper primordial en la catàlisi— tenen una participació bàsica en el manteniment d'estructures homo- o heterooligomèriques on hi figurin barrils $(\beta/\alpha)_8$ (la majoria dels barrils TIM coneguts formen part d'estructures quaternàries; Borchert et al., 1994). Un bon exemple és l'enzim triosafosfatisomerasa que sempre es troba a la natura en forma dimèrica. En aquest cas, l'interfase entre les dues subunitats està formada pels llaços que uneixen les tires β_n amb les hèlices α_n (on n agafa valors des de 1 fins a 4). Els quatre llaços contenen una gran quantitat de residus hidrofòbics que queden totalment enterrats com a resultat de la dimerització. Les operacions d'enginyeria de proteïnes que tenen com a objectiu la separació de l'homodímer de la triosafosfatisomerasa en els seus monòmers realitzen les mutacions en aquests llaços (vegeu la secciónIII.5. d'aquesta introducció).

I.4. Localització de l'activitat catalítica.

Fins avui, totes les proteïnes conegudes amb estructura de barril TIM —a excepció de la narbonina i la concanavalina B— són enzims. Els residus responsables de l'activitat enzimàtica es troben situats en la part C-terminal del barril [la majoria en els llaços que uneixen les tires β i les hèlices α (Farber i Petsko, 1990) i una quantitat menor en les pròpies tires β (Pujadas i Palau, 1997)]. Recentment, en Raychaudhuri et al. (1997) han demostrat que la cadena principal i les cadenes laterals dels barrils TIM generen un camp

electrostàtic paral·lel a l'eix del barril que apunta a una àrea propera al centre actiu. Raychaudhuri et al. (1997) han arribat a la conclusió que la presència del camp electrostàtic explicaria perquè els substrats dels barrils $(\beta/\alpha)_8$ estan majoritàriament carregats negativament.

II. Distribució dels dominis amb aquest plegament.

II.1. Distribució de les seves estructures al PDB.

El Brookhaven Protein Data Bank (BPDB o PDB; http://www.pdb.bnl.gov), és la base de dades on els cristal·lografs dipositen les coordenades de les estructures de proteïnes (6.655), àcids nucleics (530) i carbohidrats (12) (en parèntesi hi ha el nombre d'estructures depositades fins el 10 de març de 1998; Bernstein et al., 1977).

II.2. Distribució de les seves estructures al CATH.

La base de dades CATH (http://www.biochem.ucl.ac.uk/bsm/cath/; Orengo et al., 1997) és una classificació de dominis de proteïnes. Les sigles CATH corresponen als tres criteris que s'empren a l'hora de classificar un domini: classe (**C**), arquitectura (**A**), topologia (**T**) i superfamilia homologa (**H**). Els dominis amb estructura de barrils $(\beta/\alpha)_8$ estan classificats dins la classe $\alpha\beta$ ($15\leq$ % hèlice $\alpha \leq 55$ i $10\leq$ % tira $\beta \leq 45$), l'arquitectura barril i la topologia de barril TIM.

II.3. Distribució de les seves estructures al SCOP.

La base de dades SCOP (http://scop.mrc-lmb.cam.ac.uk/scop/; Murzin et al., 1995) és —a l'igual que CATH— una classificació de dominis de proteïnes. Les sigles d'SCOP signifiquen "**S**tructural **C**lasification **o**f **P**roteins". Els dominis amb estructura de barrils $(\beta/\alpha)_8$ estan classificats dins la classe α/β —caracteritzada per estar constituïda per unitats $\beta\alpha\beta$ que, a la vegada, formen una làmina β paral·lela— i dins del plegament de barril β/α o TIM —caracteritzat per la presència d'una làmina beta paral·lela en forma de barril tancat, amb 8 com a valor de **n** i **S** (vegeu definicions a la secció I.3.1.1. d'aquesta introducció) i l'ordre de les tires β és 12345678—.

II.4. Distribució en funció del seu codi enzimàtic (E.C.).

II.4.1. Distribució en funció de la seva activitat enzimàtica.

En l'actualitat, el nombre d'activitats enzimàtiques catalogades segons la classificació de la NC-IUBMB a la base de dades ENZYME (versió 22.0, novembre de 1997; http://www.expasy.ch/sprot/enzyme.html) és de 3.729 (948 oxidoreductases, 1.085 transferases, 1.072 hidrolases, 345 liases, 151 isomerases i 128 ligases). Segons les dades obtingudes a partir de la base de dades PDB_Finder [Hooft et al., (1996); versió 1.18, març de 1998; ftp://ftp.embl-heidelberg.de/pub/databases/pdb_finder], en l'actualitat s'han dipositat al Brookhaven Protein Data Bank (BPDB) (Bernstein et al., 1977) les estructures corresponents a 418 activitats enzimàtiques diferents. D'aquestes 418 activitats, 36 (un 10.8% del total) es troben en estructures amb plegament de barril TIM. Aquestes dades són coherents amb les aportades per Farber (1993) i Reardon i Farber (1995), que consideren que, aproximadament, el 10% de tots els enzims cristal·litzats tenen un domini amb estructura de barril $(\beta/\alpha)_8$. Si s'extrapolen aquests resultats es pot dir que, aproximadament, hi deu haver unes 400 activitats enzimàtiques que utilitzen com a bastida un barril TIM. Per tant, es pot concloure que la topologia de barril $(\beta/\alpha)_8$ és, de bon tros, la més comuna en els enzims i en les proteïnes globulars (Farber, 1993).

D'acord amb la classificació de la NC-IUBMB (NC-IUBMB, 1992; http://www.expasy.ch/sprot/enzyme.html), en funció de l'especificitat de substrat i del mecanisme molecular, es troben barrils TIM distribuïts en gairebé totes les classes d'enzims (8 oxidoreductases, 2 transferases, 14 hidrolases, 6 liases i 6 isomerases). La única excepció és en el cas de les ligases —on fins avui— no s'ha trobat cap barril $(\beta/\alpha)_8$.

II.4.2. El cas de la narbonina i la concanavalina B.

Fins avui, únicament dos barrils TIM —la narbonina (Hennig et al., 1992; Hennig et al., 1995b) i la concanavalina B (Hennig et al., 1995a)— no presenten cap activitat enzimàtica coneguda. Les seves estructures estan dipositades al BPDB amb els codis 1NAR i 1CNV respectivament. En ambdós casos es tracta de proteïnes d'emmagatzematge provinents de llavors de plantes leguminoses (*Vicia*

narbonensis L. en el cas de la narbonina i *Canavalia ensiformis* en el cas de la concanavalina B).

La narbonina —tot i que no se li ha detectat activitat enzimàtica— ha estat classificada per en Coulson (1994) com a una quitinasa de la família 18 de les glicosil hidrolases (Henrissat, 1991; Henrissat i Bairoch, 1993; vegeu la secció II.4.3. d'aquesta introducció). La raó cal buscar-la en la similitud significativa que existeix entre la seva seqüència i la de l'enzim endo-β-N-acetilglucosaminidasa H d'*Streptomyces plicatus* (Coulson, 1994). Aquesta homologia va fer que en Coulson (1994) hagi proposat que la narbonina té activitat quitinasa o bé deriva d'una quitinasa que ha perdut la seva funció.

Pel que respecta a la concanavalina B, la seva seqüència presenta un grau d'homologia elevat —al voltant del 40%— amb la d'altres quitinases de plantes (Hennig et al., 1995a). No obstant —i a l'igual que la narbonina— la concanavalina B no presenta activitat quitinasa. Si es compara el centre actiu de les quitinases amb la seva regió equivalent en la concanavalina B, s'observen un parell de diferències significatives (Hennig et al., 1995a). La primera consisteix en el canvi del Glu del centre actiu per una Gln i la segona en que la cavitat on s'hauria d''unir el substrat a la concanavalina B és menor que no pas la que hi ha a les quitinases. Tot i haver perdut la seva funció enzimàtica, la concanavalina B encara conserva la capacitat d'unir-se a carbohidrats. Per tant, es pot afirmar que la concanavalina B i les quitinases de la família 18 provenen d'un ancessor comú.

Estudis posteriors realitzats per Terwisscha van Scheltinga et al. (1996) han arribat a la conclusió que les quitinases de la família 18 requereixen per a la seva activitat un Asp i un Glu. A més, les seves posicions al centre actiu han de ser equivalents a les de l'Asp_{125} i del Glu_{127} en l'hevamina. La seva absència seria doncs la responsable que ni la narbonina ni la concanavalina B presentin activitat quitinasa.

II.4.3. Distribució de les glicosilhidrolases amb plegament de barril TIM en les families d'Henrissat.

És important remarcar que, de les 14 estructures amb activitat hidrolasa (vegeu la secció II.4.1. d'aquesta introducció), 12 són glicosidases amb activitat sobre substrats O-glicosídics (i per tant són O-

ISBN:978-84-694-2163-9/D.L.T.885-2011

glicosilhidrolases o més comunament glicosilhidrolases). Aquests enzims es caracteritzen perquè hidrolitzen l'enllaç glicosídic, bé entre dos o més carbohidrats, bé entre un carbohidrat i una altra molècula. El mecanisme d'hidròlisi es basa en una catàlisi àcida que requereix essencialment dos residus, un que actua com a donador de protons i un altre que fa de nucleòfil o base (Davies i Henrissat, 1995). La configuració del carboni anomèric del producte resultant de la hidròlisi pot ser, o no, igual que la del substrat emprat.

Les 12 activitats enzimàtiques EC 3.2.1.x corresponen a: α-amilasa (3.2.1.1); β-amilasa (3.2.1.2); cel·lulasa (3.2.1.4); endo-1,4-β-xilanasa (3.2.1.8); quitinasa (3.2.1.14); lisozima (3.2.1.17); β-glicosidasa (3.2.1.21); glicà endo-1,3-β-D-glicosidasa (3.2.1.39); glicà-1,4-α-maltotetrahidrolasa (3.2.1.60); liqueninasa (3.2.1.73); cel·lulosa-1,4-β-cel·lobiosidasa (3.2.1.91); manosilglicoproteïna endo-β-N-acetilglucosamidasa (3.2.1.96).

A l'inici d'aquesta dècada (Henrissat, 1991), es va proposar una classificació de les glicosilhidrolases. El fonament d'aquesta classificació rau en el fet que, en els casos que és possible, es pot establir una certa relació entre les similituds en la seqüència i el plegament corresponent. Segons la classificació d'Henrissat, les seqüències amb homologia de seqüència i homologia estructural haurien de quedar agrupades dins d'una mateixa família. Aquesta classificació presenta avantatges respecte a les basades únicament en l'especificitat de substrat [com ara la de la NC-IUBMB (NC-IUBMB, 1992)]. Per exemple, reflecteix les similituds estructurals, revela les seves relacions evolutives i proporciona un eïna a l'hora d'interpretar els mecanismes catalítics (Henrissat, 1991; Henrissat i Bairoch, 1993; Henrissat i Bairoch, 1996). Fent ús d'aquest criteri, al voltant d'unes mil seqüències —998— de glicosilhidrolases procedents de la versió 35.0 del SwissProt (març 1998) quedarien agrupades en 66 famílies (http://www.expasy.ch/cgi-bin/lists?glycosid.txt).

En la classificació de glicosilhidrolases de Henrissat (1991) els 12 enzims amb estructura barril TIM queden classificats dins de les següents famílies: família 1: β-glicosidasa; família 6: cel·lulases de la família B (cel·lulasa pròpiament); família 10: cel·lulases de la família F (endo-1,4-β-xilanasa i cel·lulosa 1,4-β-cel·lobiosidasa); família 13: α-

amilasa i glucà-1,4-α-maltotetrahidrolasa); família 14: β-amilasa; família 17: glucà endo-1,3-β-D-glicosidasa i liqueninasa; família 18: quitinases de la classe II (quitinasa pròpiament i lisozima); família 18: manosilglicoproteïna endo-β-N-acetilglucosamidasa.

III. L'enginyeria de proteïnes en els dominis amb plegament $(\beta/\alpha)_8$.

Les proteïnes amb plegament de barril TIM són estructures extraordinàriament estables. Això fa que siguin especialment adients per fer assajos sobre modulació de l'activitat enzimàtica (Bainbridge et al., 1995), obtenció d'estructures quimèriques (Mainfroid et al., 1993; Kishan et al., 1994), estabilització estructural en afegir ponts disulfur (Eder i Wilmanns, 1992) o en proporcionar hèlices α més consolidades (Mainfroid et al., 1996), establiment de dissenys *de novo* que afavoreixin aquest plegament (Richardson et al., 1987; Goraj et al., 1990; Tanaka et al., 1991; Tanaka et al., 1994a,b,c; Houbrechts et al., 1995) i, finalment, modelització 3-D a partir de variacions en la seqüència (Luger et al., 1989). Els resultats que es descriuen en aquest capítol indiquen que el plegament dels barrils $(\beta/\alpha)_8$ és extraordinàriament resistent a un ampli ventall d'operacions d'enginyeria de proteïnes, la qual cosa ha permès afirmar a Farber (1993) que "hom pot fer amb aquestes estructures gairebé qualsevol cosa i tenir encara un barril α/β". Seguidament, es descriuen les principals operacions d'enginyeria de proteïnes realitzades sobre proteïnes amb plegament de barril TIM, així com els resultats i conclusions més notables.

III.1. Permutacions circulars.

Hi han permutacions circulars naturals que s'obtenen per transposició genètica d'una peça d'ADN. En la soca salvatge, aquesta peça està col·locada a l'inici de la zona codificadora d'el gen; en canvi, en el mutant, la peça transposada està col·locada al final de la zona codificadora. Malgrat la transposició, el mutant manté la pauta de lectura. Conseqüenment, la proteïna que resulta té una estructura primària formada per dos elements de seqüència permutats respecte a altres proteïnes homòlogues.

La permutació circular, considerada com una operació d'enginyeria de proteïnes, consisteix a fer una construcció gènica que, al expressar-se en un sistema heteròleg, dóna lloc a una cadena polipeptídica permutada respecte a una proteïna natural determinada. L'operació pren el nom de permutació circular perquè fa l'efecte com si s'haguessin unit els extrems N- i C-terminals de la cadena polipeptídica natural i posteriorment "tallat" la

seqüència circular resultant en una posició diferent de l'original. Òbviament, es genera una altra estructura primària. És important elegir bé la posició de tall, donat que ha d'estar localitzada, necessàriament en un llaç. Els permutants circulars, obtinguts per tècniques d'enginyeria de proteïnes, aporten informació sobre la manera en que els extrems N- i C-terminals d'una cadena polipeptídica i també la disposició de les estructures secundàries poden afectar el plegament i l'estabilitat de les proteïnes.

Les permutacions circulars no es poden realitzar sobre qualsevol tipus de plegament, sinó únicament en aquells cassos que reuneixin dos requisits indispensables: a) que es tracti d'estructures d'una gran simetria; i b) que els extrems N- i C-terminal de la cadena polipeptídica que es vol permutar es trobin molt pròxims a l'espai. Pel que fa a la segona característica, és evident que la proximitat espacial dels dos extrems afavoreix la formació d'un enllaç peptídic entre els dos residus terminals compatible amb una mateixa estructura final, però permutada pel que fa als extrems N- i C-terminal (Luger et al., 1989). És evident que les proteïnes amb un plegament de barril $(\beta/\alpha)_8$ són una excel·lent matèria primera per a aquest tipus d'experiments degut a que compleixen, amb escreix, els dos pre-requisits esmentats. S'han identificat permutants circulars en seqüències que adopten un plegament de barril TIM (Jia et. al., 1996; MacGregor et al., 1996; Rojas i Romeu, 1996; Essen et al., 1997). Recentment, el nostre grup ha observat en uns bacteris del gènere *Clostridium* la presència d'una permutació circular en β-glucosidases de la familia 3 que afecta un domini predit com un plegament tipus barril $(\beta/\alpha)_8$ (Garcia-Vallvé et al., 1998). En aquest cas, el domini corresponent al barril TIM queda interromput per un domini tot-β situat en el llaç que uneix les unitats βα segona i tercera.

Luger et al. (1989) van realitzar un experiment pioner molt notable amb un plegament de barril TIM. Aquest experiment consistia en la construcció, expressió, purificació i anàlisi de les característiques estructurals de dos permutants de l'enzim fosforibosilantranilatisomerasa de *Saccharomyces cerevisiae* (yPRAI). En el cas particular d'yPRAI, la distància entre els extrems va aconsellar afegir cinc residus extres a l'extrem amino terminal original que facilitessin l'apropament de les posicions N- i C-terminals. La seqüència resultant d'aquesta addició va ser anomenada **x**PRAI (que deriva de yPRAI extesa). Les anàlisis d'activitat enzimàtica van demostrar que les propietats catalítiques d'yPRAI i d'xPRAI eren molt similars.

Consequentment, es va concloure que xPRAI era una excel·lent matèria primera sobre la qual poder realitzar permutacions circulars.

III.2. Disseny de novo.

El disseny *de novo* de proteïnes és una metodologia relativament nova (DeGrado, 1988; DeGrado et al., 1989; Richardson i Richardson, 1989; Cohen i Parry, 1990; Hecht et al., 1990; Sander et al., 1992) que té com objectiu la proposta de cadenes polipeptídiques el plegament de les quals corresponguin a una estructura 3-D predeterminada. La denominació *de novo* fa referència a que les seqüències que es proposen han de ser significativament diferents a qualsevol altra seqüència natural coneguda. Aquesta metodologia utilitza tot un calidoscopi de tècniques que van des del modelatge de l'estructura en una estació gràfica fins al clonatge i expressió del gen, la purificació de la proteïna i la determinació posterior de les característiques estructurals pròpies.

La gran quantitat d'estructures amb plegament de barril $(\beta/\alpha)_8$ i seqüència no homòloga (vegeu la secció II d'aquesta introducció) fa que la bastida estructural sigui especialment apropiada per "construir-hi" noves funcions (Lasters et al., 1988) o emprar-la en un exercici de disseny *de novo* (Richardson et al., 1987; Goraj et al., 1990; Tanaka et al., 1991; Tanaka et al., 1994a,b,c; Houbrechts et al., 1995). En aquest sentit, no és d'estranyar que el primer disseny d'un barril TIM es fes ja l'any 1986, durant un curs que es va realitzar al Laboratori de Biologia Molecular Europeu [EMBL; Heidelberg, Alemanya (Richardson et al., 1987)]. La proteïna dissenyada en aquell curs va rebre el nom de *Babarellina* i es va construir a partir de quatre repeticions d'una unitat bàsica βαβα. La longitud de la molècula era de 178 residus. Un cop fet el disseny de la seqüencia, es va sintetizar el gen que l'havia d'expressar i es va clonar, tot i que sense obtenir-se'n cap resultat satisfactori (Dr. Chris Sander, comunicació personal).

Posteriorment, dos grups de recerca van realitzar, amb més o menys èxit, el disseny i la síntesi*de novo* d'un barril TIM. El grup liderat pel Dr. Joseph Martial de la Universitat de Lieja (Bèlgica) va obtenir una cadena polipeptídica que anomenaren Octarellina (Goraj et al., 1990). La seqüència de l'Octarellina està formada per una unitat bàsica de 31 residus (SGLVVYLGKR PDSGTARELL RHLVAEGDAR S) que es repeteix vuit vegades. En aplicar el mètode de GOR (Garnier et al., 1978), l'estructura secundària predita per a aquest segment repetitiu era del tipus βα. La proteïna es va

expressar en *E. coli* i i, després de ser purificada, es va sotmetre a una anàlisi estructural mitjançant tècniques de dicroisme circular, electroforesi amb gradient d'urea i espectroscòpia d'infraroig i Raman (Beauregard et al., 1991). La principal conclusió derivada de les anàlisis fou que la molècula, tot i presentar hèlices α i tires β força estables, no tenia un bon empaquetament i, per tant, calia millorar-ne el disseny. El redisseny fou efectuat per Houbrechts et al. (1995). En aquest treball, es van tenir en compte diversos aspectes que originalment no havien estat considerats (distribució de càrregues, volum i hidrofobicitat dels residus propis de les làmines β i la simetria de la seqüència). En certa manera, van recollir les suggerències que el Prof. Chris Sander els havia fet en una revisió publicada l'any 1991 (Sander, 1991). En aquest article, Sander —basant-se en l'anàlisi dels barrils TIM existents a la natura— suggeria que era més correcte repetir quatre vegades una mateixa unitat βαβα que no pas vuit vegades una unitat βα. A més, aconsellava disminuir l'acumulació de càrregues (16 en total) en un dels extrems del barril de l'Octarellina la qual cosa afavoriria la formació de l'estructura. Tenint en compte les suggerències de Sander (1991), Houbrechts et al. (1995) van dissenyar dues cadenes polipeptídiques noves que anomenaren Octarellina II i III. La seqüència II està formada per vuit repeticions de la unitat $\alpha\beta_1$ i la III per quatre repeticions de $\alpha\beta_1\alpha\beta_2$. Per tant, l'única diferència entre elles era que en la III hi havia dos tipus de tires β diferents, β_1 i β_2, mentres que en la II totes eren del tipus β_1. Un cop expressades i purificades Es va demostrar que l'Octarellina III purificada era força estable a la desnaturalització amb urea. Aparentment, aquest resultat corrobora la importància de dissenyar aquest tipus de molècules a partir de quatre repeticions d'una unitat αβαβ', tal com havia suggerit Sander (1991).

Un segon grup de recerca liderat per Toshiki Tanaka i Haruki Nakamura (PERI, Protein Engineering Research Institute; Osaka, Japó) també ha estat treballant en el camp del disseny *de novo* de proteïnes amb plegament de barril TIM (Tanaka et al., 1991; Tanaka et al., 1994a,b,c). En un primer treball (Tanaka et al., 1991) van dissenyar quatre barrils $(\beta/\alpha)_8$ de 217 residus cadascun a partir d'una unitat bàsica $\beta_1\alpha_1\beta_2\alpha_2$ repetida quatre cops. Després de ser expressades i purificades, les quatre proteïnes es van desnaturalitzar amb urea. Això va permetre identificar l'estructura més estable que va rebre el nom de TIM-1. La unitat bàsica $\beta_1\alpha_1\beta_2\alpha_2$ de TIM-1 va ser utilitzada per a dissenyar dos "barrils" nous, el $(\beta/\alpha)_6$ i el $(\beta/\alpha)_{10}$. Finalment, es va comparar l'estabilitat de TIM-1, $(\beta/\alpha)_6$ i $(\beta/\alpha)_{10}$. Es va concloure que la proteïna artificial $(\beta/\alpha)_{10}$ era la més estable, tant com ho és

ISBN:978-84-694-2163-5/DL:T.885-2011

la proteïna TIM natural. Tanaka et al. (1991) van justificar aquests resultats adduint que tant les interaccions com l'empaquetament entre cadenes laterals a l'interior del barril eren més afavorits en el cas de $(\beta/\alpha)_{10}$ que no pas en el cas de TIM-1. En conseqüència, aquests autors van arribar a la conclusió que, per al disseny *de novo* d'una estructura $(\beta/\alpha)_8$, probablement és crucial fer una bona elecció dels residus que han de formar l'interior del barril. Així, per fer la tria de residus, caldrà coneixer les característiques bàsiques (volum de la cadena lateral, hidrofobicitat, etc.) de les combinacions naturals de residus en aquestes posicions. Segons Tanaka et al. (1991), el més important és combinar cadenes laterals que, no només tinguin facilitat d'interaccionar entre elles, sinó que també tinguin uns tamanys complementaris (per exemple, compensant els volums de les cadenes laterals). Procedint d'aquesta manera, si es divideix el barril de TIM-1 en seccions equatorials que passen per quatre residus de tires β alternades (i amb les cadenes laterals encarades cap a l'interior del barril), s'observa que cadascuna de les seccions conté quatre residus idèntics. A partir d'aquestes consideracions, Tanaka et al. (1991) van apuntar que la millora del disseny de l'estructura de TIM-1 hauria de passar ineludiblement per la correcció dels defectes d'empaquetament causats per la confluència de residus idèntics.

Posteriorment, Tanaka et al. (1994a) van fer un nou disseny d'un barril TIM a partir d'una unitat bàsica $\beta_1\alpha_1\beta_2\alpha_2$ que es repetia quatre vegades. En aquest cas —a diferència del disseny realitzat el 1991 (Tanaka et al., 1991)— es va tenir cura de combinar el tamany de les cadenes laterals que havien de quedar dins del barril per tal que poguessin cabre en l'espai que deixen les vuit tires β al seu interior. Es va comprovar per microscòpia electrònica que la proteïna resultant tenia una forma similar a la de l'enzim triosafosfatisomerasa de conill. Tot i aquest darrer resultat, els autors van reconeixer que només es pot tenir la certesa que el plegament de la proteïna dissenyada *de novo* és del tipus TIM si s'aconsegueix realitzar una determinació estructural més precisa (mitjançant raigs-X o RMN).

Aquest grup ha dissenyat, amb més o menys èxit, altres cadenes polipeptídiques les seqüències de les quals podrien adquirir aquest tipus de plegament (Tanaka et al., 1994b,c). Un dels millors exemples és el disseny realitzat per Tanaka et al. (1994c) a partir de la duplicació d'una unitat bàsica $\beta_1\alpha_1\beta_2\alpha_2\beta_3\alpha_3\beta_4\alpha_4$. En aquest cas, la seqüència es va dissenyar de forma molt acurada, buscant la compatibilitat amb totes aquelles característiques generals en les estructures amb plegament de barril $(\beta/\alpha)_8$ (empaquetament

ISBN:978-84-694-2163-5/DL:T.885-2011

òptim de les cadenes laterals a l'interior del barril; disseny d'hèlices amfipàtiques que haurien de facilitar la formació de ponts salins intrahelicoidals; disseny de llaços curts a l'extrem N-terminal). Els llaços de l'extrem C-terminal del barril es van manllevar de proteïnes amb plegament TIM. Posteriorment, es va construir un model de la molècula al qual es va aplicar una minimització energètica. Pel que fa a l'estructura secundària, es va comprovar que hi havia un grau molt elevat de coincidència entre el disseny i la predicció feta per diferents mètodes. També es van fer anàlisis de compatibilitat entre la seqüència i el plegament 3-D propi d'un barril TIM. Aquests tests semblaven indicar que l'estructura final modelada hauria de tenir unes característiques similars a les proteïnes naturals amb plegament de barril $(\beta/\alpha)_8$. Després de ser expressada i purificada, la proteïna presentava un empaquetament deficient que era compatible, però, amb la presència de proporcions ajustades d'estructura secundària, un nucli hidrofòbic característic i un plegament cooperatiu.

Finalment, cal esmentar el grup de la Dra. Wodak (Universitat Lliure de Brusel·les, Bèlgica) va estar treballar durant un temps en la parametrització de les característiques geomètriques dels barrils $(\beta/\alpha)_8$ (vegeu la secció I.3.1.1. d'aquesta Introducció). Segons els autors, aquesta parametrització hauria de sentar les bases teòriques per al disseny *de novo* d'un barril TIM (Lasters, 1990; Lasters et al., 1990; Wodak et al., 1990).

III.3. Disseny de barrils $(\beta/\alpha)_n$ ($n \neq 8$).

Els dominis amb estructura de barril TIM han estat utilitzats com a matèria primera per al disseny de barrils $(\beta/\alpha)_n$. Les estructures en forma de barril TIM emprades provenen de proteïnes naturals (Luger et al., 1990) o bé de dissenys *de novo* que, suposadament, presenten aquest plegament (Tanaka et al., 1991). Els barrils $(\beta/\alpha)_n$ ($n \neq 8$) dissenyats fins ara són $(\beta/\alpha)_6$ (Tanaka et al., 1991) i $(\beta/\alpha)_{10}$ (Luger et al., 1990; Tanaka et al., 1991). L'experiment realitzat per Luger et al. (1990) va consistir a afegir a l'extrem N-terminal del gen que codifica l'enzim yPRAI el fragment d'ADN corresponent al segment $\alpha_7\beta_7\alpha_8\beta_8$ de la proteïna. El resultat d'aquesta operació va ser un gen que codificava per una proteïna del tipus $\alpha_7\beta_7\alpha_8\beta_8\alpha_1\beta_1.\alpha_2\beta_2...\alpha_7\beta_7\alpha_8\beta_8$. El gen es va expressar posteriorment en *E. coli*. Els detalls experimentals sobre estabilitat estructural activitat i enzimàtica estan descrits als treballs originals (Luger et al., 1990; Tanaka et al., 1991), els quals suggereixen que els barrils $(\beta/\alpha)_{10}$ presenten característiques similars als barrils TIM dels quals deriven.

ISBN:978-84-694-2162-5/DL:T-885-2011

III.4. Disseny de ponts disulfur extres.

Un pont disulfur consisteix en un enllaç covalent entre dos àtoms de sofre corresponents a les cadenes laterals de dos residus de cisteïna. La formació d'aquest enllaç covalent produeix una oxidació dels dos grups tiol que hi participen. Al conjunt de les dues cisteïnes unides covalentment s'anomena cistina. Normalment, la formació de ponts disulfur i el plegament de la cadena polipeptídica són processos simultanis (Darby i Creighton, 1995; Raina i Missiakas, 1997). A més de facilitar el plegament correcte, els ponts disofre participen en el manteniment de l'estructura tridimensional i augmenten la termoestabilitat de la proteïna (Wetzel 1987; Creighton, 1988). El mecanisme pel qual els ponts disulfur estabilitzen una estructura sembla estar relacionat amb la disminució d'entropia que duen associada (Anfinsen i Scheraga, 1975; Chan i Dill 1989; Matsumura et al., 1989b). Per tant, no és d'estranyar que diversos autors s'hagin interesat per introduir ponts disulfur extres mitjançant tècniques d'enginyeria de proteïnes (Perry i Wetzel, 1984; Sauer et al., 1986; Villafranca et al., 1987; Matsumura et al., 1989a,b; Mitchunson i Wells, 1989; Kanaya et al., 1991).

Els barrils TIM també poden augmentar la termoestabilitat mitjançant l'addició de ponts disulfur extres. Eder i Wilmanns (1992) van publicar el primer —i fins ara únic— treball sobre la creació d'una nova cistina en un barril $(\beta/\alpha)_8$. La proteïna emprada era el mutant $Ser^{212}Cys$ d'yPRAI —que en la seva forma natural no conté cap pont disulfur— i el nou enllaç covalent es va crear entre els S^{γ} de la Cys_{27} i la Cys_{212}. Aquesta cistina uneix dues hèlices adjacents de l'estructura (α_1 i α_8) i forma una proteïna quasi-circular. A més, s'ha demostrat que el pont disulfur es forma espontàniament *in vitro* i estabilitza la conformació nativa d'yPRAI en 1.0 Kcal/mol (Eder i Wilmanns, 1992).

III.5. Transformació d'oligòmers de barrils$(\beta/\alpha)_8$ en monòmers.

La majoria dels barrils $(\beta/\alpha)_8$ coneguts formen part d'homo- o heterooligòmers (Borchert, 1994). En molts casos l'oligòmer juga un paper fonamental en que l'enzim tingui una estabilitat i una activitat catalítica superior a la que li correspondria si fós un monòmer. Un bon exemple seria el de l'enzim triosafosfatisomerasa —que dóna nom al plegament— i que, per regla general, es troba en forma dimèrica. Per tant, un objectiu científic molt interessant consisteix a coneixer l'aportació de cada monómer a l'activitat total de l'enzim.

ISBN:978-84-694-2163-5/DL:T.885-2011

El grup liderat pel Dr. Wierenga al Laboratori de Biologia Molecular Europeu (EMBL; Heidelberg, Alemanya) va plantejar-se aquest objectiu (Borchert et al., 1993; Borchert et al., 1994). Els autors van escollir l'enzim triosafosfatisomerasa de *Trypanosoma brucei* per realitzar els seus experiments. El dímer d'aquest enzim té una interfase de contacte molt extensa que implica una superfície de 1.600 $Å^2$ per subunitat. Bàsicament, la interfase està formada pels llaços que uneixen les tires β_n amb les hèlices α_n (on n agafa valors des de 1 fins a 4). El centre actiu de cada subunitat es troba prop de la interfase. No obstant això, els residus d'una subunitat no interaccionen directament amb el substrat que es troba al centre actiu de l'altra. El llaç L_3 que uneix β_3 i α_3 aporta a l'interfase del dímer les interaccions estabilitzants més importants (molts dels residus implicats es troben enterrats a la interfase). En conseqüència, el llaç L_3 va ser escollit per fer les modificacions que haurien de convertir el dímer en monòmer. Al tractar-se d'un llaç fortament hidrofòbic, les modificacions s´havien de fer amb molta cura, i no n'hi havia prou en fer mutacions puntuals que evitessin la formació del dímer. A més, calia impedir que aquesta part de la molècula —d'alt contingut en residus hidrofòbics— quedés exposada al solvent i endegués canvis conformacionals que podrien alterar de forma significativa l'estructura de la proteïna. Totes aquestes consideracions van ser tingudes en compte en el model que els autors van realitzar abans de prendre una decisió sobre el mutant que que calia expressar. Els resultats del modelatge van aconsellar de substituir un segment de 15 residus a L_3 —del 68 fins al 82— per un de vuit. Una vegada dissenyada la seqüència d'aminoàcids corresponent al monòmer, el seu gen es va expressar en *E. coli* i la proteïna resultant va ser purificada i cristal·litzada. L'anàlisi de l'estructura va demostrar que l'enzim mantenia les característiques estructurals corresponents a un barril TIM. A més, conservava part de l'activitat catalítica del dímer (la seva k_{cat} era 1000 vegades menor i la K_m augmentava per un factor de 20 en relació als valors propis de l'enzim natural). El monomer resultant d'aquest experiment va rebre el nom de "monoTIM".

III.6. Disseny de proteïnes quimèriques amb estructura de barril $(\beta/\alpha)_8$.

Les investigacions per esbrinar el nivell d'estabilitat i la capacitat de plegament de proteïnes, l'estructura de les quals està formada per barrils TIM, han utilitzat mutants que presenten canvis puntuals respecte a l'enzim natural (Yutani et al., 1980; Borchert et al., 1993; Tsuji et al., 1993). Una manera més dràstica d'afrontar el mateix problema consisteix a canviar

segments sencers de la seqüència (estructures secundàries, supersecundàries, motius en general, etc.) i comparar les característiques estructurals i catalítiques amb les de l'enzim original. Al fer aquests canvis cal tenir cura de minimitzar l'impacte sobre la resta de l'estructura. Una manera d'aconseguir-ho consisteix a substituir el segment de la seqüència original per un d'equivalent del mateix enzim però de font diferent. La seqüència obtinguda amb aquest procediment s'anomena "quimèrica" o "híbrida".

Com a matèria primera per a la construcció de "quimeres" s'han emprat barrils $(\beta/\alpha)_8$ (Mainfroid et al., 1993; Kishan et al., 1994). L'objectiu d'aquests estudis consistia a determinar l'efecte d'un canvi d'aquestes característiques sobre l'estabilitat i l'activitat de l'enzim. Des del punt de vista estructural, els barrils TIM poden considerar-se pseudosimètrics. Aquesta pseudosimetria es basa en l'existència de vuit motius consecutius del tipus llaç–tira β–llaç–hèlice α (L-β-L-α a partir d'ara) que poden considerar-se com a unitats de plegament. Aquest visió dels barrils $(\beta/\alpha)_8$ serveix de justificació teòrica als experiments que es plantegin construir un híbrid de barril TIM mitjançant la substitució d'un motiu L-β-L-α.

L'experiment realitzat per Mainfroid et al. (1993) consistia a substituir el vuité motiu L-β-L-α [$_8$(L-β-L-α) a partir d'ara] de l'enzim triosafosfatisomerasa de *E. coli* pel seu equivalent en el pollastre. Aquest motiu té una longitut de 23 residus. La proteïna híbrida resultant va rebre el nom d'ETIM8CHI. La principal raó per a escollir $_8$(L-β-L-α) és que no participa en les interaccions entre subunitats (vegeu la secció III.5.). A més, $_8$(L-β-L-α) està involucrat en la unió del substrat al centre actiu. Així, el residu Leu_{233} es troba a la base del centre actiu i els residus Gly_{235} i Gly_{236} són a l'extrem N-terminal d'una hèlice 3_{10} que té una participació important en la unió del grup fosfat del substrat amb l'enzim. La creació d'un híbrid que empra el motiu on hi han els residus Leu_{233}, Gly_{235} i Gly_{236} —totalment conservats en un nombre molt gran de seqüències de triosafosfatisomerasa— serveix per a comprovar si els tres residus adopten la conformació necessària per a poder participar en la catàlisi. Per altre part, el canvi del motiu $_8$(L-β-L-α) també afecta l'empaquetament hidrofòbic entre la tira β_8 i l'hèlice α_7. Això es degut a que un dels 10 canvis que introdueix el nou motiu $_8$(L-β-L-α) és un residu Phe en comptes d'Ala en la posició 232 de la seqüència d'aminoàcids. Els resultats obtinguts mitjançant el modelatge per homologia de l'híbrid mostren que el residu Phe_{232} pot ser enquibit dins la interfase de β_8 i α_7 si, simultàniament, rota la cadena lateral de Phe_{224}. Després

ISBN:978-84-694-2163-5/DL.T.885-2011

d'expressar el producte quimèric en *E. coli* i purificar-lo, es van fer tests d'estabilitat i estudis comparatius sobre l'activitat enzimàtica. Els estudis cinètics van demostrar que l'híbrid tenia una certa activitat (la k_{cat} venia disminuida per un factor de 10 respecte a la pròpia de l'enzim TIM d'*E. coli*). A més, l'estabilitat estructural d'ETIM8CHI era semblant —tot i que no idèntica— a la de l'enzim natural d'*E. coli*.

III.7. Disseny d'alteracions en l'activitat catalítica d'enzims naturals amb estructura de barril $(\beta/\alpha)_8$.

A la secció III.2. s'han descrit les estratègies seguides per dissenyar seqüències que adoptin un plegament prefixat. L'objectiu final d'aquests estudis, que empren com a bastida una estructura de barril $(\beta/\alpha)_8$, és determinar els mecanismes que governen el plegament dels barrils TIM. El seu coneixement permetria dissenyar enzims "a la carta", és a dir, molècules termostables —inexistents a la natura— posseïdores d'activitats enzimàtiques amb una eficiència catalítica millorada.

Un altre exemple de disseny de funcions —més modest en els seus objectius que no pas els dissenys *de novo*, però no per això menys interessant— consisteix a alterar les propietats catalítiques d'un enzim. Un bon exemple és el treball realitzat per Bainbridge et al. (1995) amb l'enzim rubisco (ribulosa 1,5-bisfosfat carboxilasa/oxigenasa; EC 4.1.1.39). Aquest enzim catalitza dues reaccions: a) la carboxilació de la D-ribulosa-1,5-bisfosfat que és el pas inicial en la fixació del CO_2 mitjançant la fotosíntesi [D-ribulosa-1,5-bisfosfat + CO_2 —> 2(3-fosfo-D-glicerat)]; i b) la fragmentació oxidativa de la D-ribulosa-1,5-bisfosfat en la fotorespiració [D-ribulosa-1,5-bisfosfat + O_2 —> 3-fosfo-D-glicerat + 2-fosfoglicolat]. Les dues reaccions són simultànies, es produeixen en el mateix centre actiu i estan en competència. L'enzim és un oligòmer i la seva composició depen de si prové de gens de la forma I o de la forma II. Els que provenen de la forma I estan constituïts per 8 còpies de dues cadenes polipeptídiques diferents: la gran o L (de l'anglès *large*) i la petita o S (de l'anglès *small*). Per la seva part, els que estan codificats dins l'operò de la forma II estan formats únicament per subunitats grans. El centre actiu està localitzat a les subunitats grans (http://www.biochem.ucl.ac.uk/cgi-bin/cath/CATHSrch.pl?type=GEN&query=rubisco).

L'objectiu de l'experiment realitzat per Bainbridge et al. (1995) consistia a buscar els canvis que cal fer en la seqüència de l'enzim per afavorir l'activitat assimiladora de carboni respecte a la consumidora d'oxigen. La modulació

de l'activitat enzimàtica es va estudiar emprant la subunitat gran d'*Anacystis nidulans*. Les mutacions es van realitzar en el segment de gen que codifica la seqüència de l'extrem C-terminal d'L_6 [llaç que uneix a β_6 i α_6 en el barril $(\beta/\alpha)_8$]. Es va escollir L_6 perquè influeix sobre l'especificitat de l'enzim pels seus substrats gasosos CO_2 i O_2, per bé que interacciona amb parts de la molècula diferents de les del centre catalític. Un exemple característic és el residu Lys_{128} que es troba aprop tant del substrat com de diversos grups carbonil que pertanyen a la cadena principal de residus d'L_6. Bainbridge et al. (1995) van comprovar que la mutació de Lys_{128} a residus de menor basicitat reduia l'activitat enzimàtica i, a més, produia l'efecte contrari al desitjat, és a dir, afavoria l'oxigenació respecte la carboxilació.

IV. Evolució globalitzada *versus* evolució individual dels barrils TIM: divergència o convergència evolutives.

La possible relació evolutiva entre els diferents dominis estructurals tipus TIM ha estat motiu d'estudi des de la primera meitat de la dècada dels vuitanta. Muirhead (1983) ja va indicar que la presència d'aquest plegament en diversos enzims —aparentment no relacionats pel que fa a l'estructura primària— podia ser fruit d'un procés evolutiu bé convergent (seqüències no relacionades que convergeixen en un plegament tridimensional comú) o bé divergent (evolució de les seqüències a partir d'un avantpassat comú). Posteriorment, a partir de l'observació de la posició dels introns en el gen que codifica l'enzim piruvatkinasa, Lonberg i Gilbert (1985) van proposar que els barrils TIM podrien haver-se format mitjançant combinació d'exons. Segons les observacions que van realitzar en aquest gen, els seus introns no es disposen a l'atzar sinó que divideixen la seqüència codificant en peces que contenen elements discrets de l'estructura secundària. Més concretament, cadascun dels sis exons contribueix a l'estructura del barril amb una o dues repeticions de la unitat hèlice α – tira β. Sobre aquesta base, Lonberg i Gilbert (1985) van suggerir que els introns no estan superposats interrompent una seqüència codificant que existia amb anterioritat sinó que són producte de l'evolució del primer gen de la piruvatkinasa.

La gran quantitat de seqüències i estructures de barrils TIM obtingudes des de llavors han permés donar arguments a favor i en contra d'aquestes hipòtesis evolutives. A continuació es recullen quines són les evidències experimentals més importants que donen suport a una o a altre teoria evolutiva.

IV.1. La hipòtesi evolutiva convergent.

Lesk et al.(1989) van analitzar l'empaquetament dels residus en el nucli del barril dels enzims triosafosfatisomerasa, rubisco i glicolatoxidasa (vegeu la secció I.3.1.2. d'aquesta introducció). Els resultats van demostrar que la manera com s'organitzen els residus interns de la làmina β és diferent en la glicolatoxidasa de la que es troba en els altres dos enzims. Aquests resultats feien suggerir als autors (Lesk et al., 1989) que no totes les proteïnes amb aquest tipus de plegament han d'estar necessàriament relacionades evolutivament, sinó que constitueixen un bon exemple de solució comuna als problemes relacionats amb la creació d'un barril tancat (el plegament TIM seria molt més estable que el que presentaven inicialment).

IV.2. La hipòtesi evolutiva divergent.

Lesk et al. (1989) havien interpretat que les diferències en l'empaquetament dels residus interns del domini $(\beta/\alpha)_8$ era l'argument principal a favor d'una evolució convergent cap a aquest tipus de bastida estable. Raine et al. (1994), però, van demostrar mitjançant la utilització d'alineacions estructurals dels barrils de les flavooxidases/deshidrogenases que determinades alteracions genètiques molt simples podrien ser responsables d'aquests dos tipus d'empaquetaments diferenciats (triosafosfatisomerasa i rubisco, d'una banda, i glicolatoxidasa, de l'altra) i, en conseqüència, van concloure que els dos tipus de nuclis que es troben en els dominis amb estructura de barril TIM poden provenir perfectament d'un avantpassat comú (evolució divergent). Segons els mateixos autors (Raine et al., 1994), seria innecessària una explicació basada en una evolució convergent a l'hora de descriure aquest fenomen. Aquesta opinió ha estat recentment contestada per Janecek i Bateman (1996), els quals consideren que aquest factor, per si sol, no aclareix la qüestió de l'evolució dels barrils $(\beta/\alpha)_8$.

IV.3. Característiques comunes a diferents dominis amb plegament $(\beta/\alpha)_8$. Evidències d'una evolució divergent.

IV.3.1. Llocs d'unió a fosfats.

Una característica comuna a molts dels barrils $(\beta/\alpha)_8$ és la presència de llocs d'unió a grups fosfat en la part C-terminal del seu barril. Aquest motiu va ser descrit inicialment per Wilmanns et al. (1991) en tres enzims amb estructura de barril TIM que es troben situats en tres pases

succesives de la ruta metabòlica de la biosíntesi del triptòfan. Els enzims són el fosforribosil antranilat isomerasa, l'indol-glicerol fosfatsintasa i el triptofansintasa (subunitat α). Els llocs d'unió a grups fosfats es troben en posicions equivalents en les tres estructures i comprenen els residus del llaç que uneix la β_7 amb l'hèlice α_7 a més de la part N-terminal d'una hèlice 8' addicional. Wilmanns et al. (1991) també van assenyalar la presència de llocs d'unió a grups fosfat en d'altres sis enzims d'idèntic plegament però que no es troben relacionats funcionalment amb els que participen en la biosíntesi del triptofan. Aquests enzims són la triosafosfatisomerasa, la rubisco, el glicolatoxidasa, el flavocitocrom b2, la trimetilamina deshidrogenasa i la fructosa bisfosfat aldolasa. Més recentment, Bork et al. (1995) han detectat llocs d'unió a grups fosfat en un gran nombre de barrils TIM mitjançant la cerca del corresponent motiu. En la majoria de casos en que es va detectar aquest motiu, no es coneixia la corresponent estructura terciària. No obstant això, els mètodes de predicció d'estructura secundària van revelar una alternança entre tires β i hèlices α que estaria d'acord amb la topologia típica dels barrils $(\beta/\alpha)_8$. El posicionament del motiu en les estructures secundàries predites era també coherent amb la seva localització en les estructures terciàries (entre la tira β_7 i l'hèlice 8).

IV.3.2. Barrils depenents de flavinmononucleòtids (FMN).

Les estructures terciàries de les proteïnes glicolatoxidasa, flavocitocrom b2 i trimetilamina deshidrogenasa (Lindqvist et al., 1991) —a més del lloc d'unió a fosfat (vegeu la secció IV.3.1. d'aquesta introducció)— també presenten altres semblances. Per exemple, són depenents de flavinmononucleòtids (FMN), presenten un plegament de barril $(\beta/\alpha)_8$, el grup FMN s'uneix a ells de forma similar i el seu lloc d'unió està situat en les tres estructures en l'extrem C-terminal del barril. En el cas dels enzims glicolatoxidasa i flavocitocrom b2 —amb una homologia en les seqüències dels dominis TIM del 37%— la similitud entre les seves estructures sobrepassa el propi barril i s'exten cap fora del domini $(\beta/\alpha)_8$.

IV.3.3. Les "homologies amagades" d'Stefan Janecek.

En els darrers anys, el professor eslovac Stefan Janecek ha encaminat els seus esforços a buscar característiques comunes de proteïnes amb plegament de barril TIM, i de manera especial de glicosilhidrolases.

Entre les seves aportacions més destacades, cal remarcar la detecció de glicines i prolines situades en els llaços que hi ha al voltant de les tires β2 i β4 (Janecek, 1995; Janecek, 1996) en 17 estructures diferents. En el cas de la α-amilasa, aquests residus se situen al voltant de la tira β2 mentre que en el cas de la glicolatoxidasa són al voltant de la β4. La subunitat α de l'enzim triptòfansintasa té les seves tires β2 i β4 similars a la β2 de la α-amilasa i a la β4 de la glicolatoxidasa. Aquest tipus d'homologies han estat batejades per Janecek com "homologies amagades" (Janecek, 1996) i la seva presència estaria en consonància amb el fet que podria haver-hi hagut una llarga història evolutiva de tots els barrils $(\beta/\alpha)_8$. El mateix autor també ha descrit la presència d'un residu d'àcid glutàmic prop de l'extrem C-terminal de la tira β5 (Janecek i Balaz, 1995). Aquest residu va ser trobat en 12 barrils TIM diferents que corresponen a 3 famílies estructurals diferents i juga un paper molt important en la funcionalitat d'aquests enzims.

IV.3.4. L'evolució divergent global dels barrils TIM segons Farber i Petsko.

Una altra via d'estudi de l'evolució dels barrils TIM que ha obtingut molt de ressó ha estat la seguida per Farber i Petsko (1990). Aquests autors van tractar d'estudiar les relacions evolutives que hi havia entre els enzims amb estructura de barril TIM coneguts fins les hores (17 concretament). Per tal de realitzar-ho van classificar les diferents estructures en quatre famílies i dins d'aquestes, i únicament quan fos necessari, en sotsfamílies. Aquestes famílies es van constituir a partir de l'ànalisi de les estructures cristal·litzades. Més concretament, les característiques estructurals analitzades van ser: la longitud de les hèlices i les tires β que constitueixen el barril, el nombre i la localització de dominis, d'hèlices o de tires β extres i la localització de l'eix major de l'elipse del barril. Aquests autors van constatar així mateix una bona correlació entre les famílies (basades únicament en les característiques estructurals abans esmentades) i o bé la participació del seus membres en vies metabòliques comunes o bé en la dependències de cofactors comuns. A nivell d'homologia entre les seqüències, aquests autors no van observar cap homologia entre enzims que pertanyessin a diferents famílies. La principal conclusió a la qual van arribar va ser que totes aquelles 17 estructures provenien d'un mateix avantpassat mitjançant divergència evolutiva. Els arguments proposats per tal d'explicar

aquesta evolució es basaven principalment en el fet que totes les estructures tenien activitat enzimàtica i que la localització del centre actiu a l'extrem C-terminal del barril (en els llaços que connecten les tires β i les hèlices α) era comú a tots elles. A partir de la línia marcada per aquest treball, es van desenvolupat els estudis posteriorment realitzats per Farber (1993) i Reardon i Farber (1995). En aquests treballs es va procedir a fer una actualització de l'estudi realitzat per Farber i Petsko (1990) incorporant-hi aquells barrils que s'havien descobert amb posterioritat a aquest estudi inicial. Com a resultat d'aquesta actualització, i seguint la mateixa metodologia que en el treball original, es van incorporar dues noves famílies a les quatre que ja s'havien descrit previament.

IV.3.5. L'evolució de la superfamília de les glicosilhidrolases.

Les glicosilhidrolases amb estructura de barril TIM, i especialment aquelles involucrades en la degradació del midó, també han estat estudiades extensament des del punt de vista evolutiu. Atès que les grans diferències pel que fa a l'estructura primària entre aquests enzims fan inviable un estudi evolutiu que empri la totalitat de la seqüència, molts treballs s'han centrat en la identificació dels residus involucrats en la catàlisi. La distribució de les glicosilhidrolases amb plegament de barril TIM en les families d'Henrissat s'ha vist a la secció II.4.3. d'aquesta introducció.

Henrissat et al. (1995) van analitzar la presència de motius comuns en les regions que envolten els residus catalítics coneguts d'algunes O-glicosilhidrolases. El mateix estudi va demostrar que aquests motius també es troben en d'altres 150 seqüències de glicosilhidrolases. Aquest grup de seqüències, molt heterogeni, està distribuït en 8 famílies diferents (Henrissat, 1991; Henrissat i Bairoch, 1993) i actuen sobre una gran varietat de substrats. Dues famílies mantenen una estructura en forma de barril TIM i s'ha suggerit que les altres 6 famílies també poden compartir aquest plegament en el seu domini catalític. Així doncs, les glicosilhidrolases il·lustren la manera en què els esdeveniments evolutius han menat a obtenir una diversificació d'especificitat de substrat amplia, amb una disposició dels residus catalítics idèntica a partir possiblement del mateix ancestre (*ancestor* en anglès) amb característiques estructurals tipus TIM.

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8:
RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

Objectius.

La secció IV. d'aquesta introducció reflecteix la controversia que existeix a l'entorn de l'evolució dels barrils TIM. Normalment, els estudis evolutius han partit d'un calidoscopi d'estructures i seqüències que pertanyen a diferents proteïnes i han intentat arribar a teories evolutives globalitzadores (Farber i Petsko, 1990; Janecek, 1994; Janecek, 1995). Aquest tipus d'estudis però, presenten inconvenients metodològics importants. El principal rau en la gran diversitat que —en funció de l'origen del ser viu— presenten les seqüències d'una mateixa proteïna. En aquest cas, és coherent agrupar les seqüències en famílies si l'única font d'informació que es considera és la proporcionada per les estructures cristal·litzades (Farber i Petsko, 1990)?. Sembla evident que la resposta és que no. Per tant, sembla necessari desenvolupar una metodologia que permeti incorporar en els estudis evolutius dels barrils TIM la informació que proporcionen les seqüències de proteïnes —encara— no cristal·litzades.

La proposta metodològica que es presenta en aquesta tesi es basa en la necessitat de fer un estudi evolutiu molt acurat —i individualitzat— de cadascuna de les proteïnes que presenten un plegament de barril TIM. Aquest estudi ha de marcar-se com a objectiu la detecció d'aquelles característiques que —suposadament— han de tenir en comú i que han d'estar relacionades amb l'activitat enzimàtica i l'estabilitat estructural. A l'hora de desenvolupar aquesta metodologia cal escollir com a model una proteïna amb plegament de barril $(\beta/\alpha)_8$. La proteïna que s'ha escollit com a model és la β-amilasa. Les raons de la seva elecció cal buscar-les en que es coneix la seva estructura —tant en forma lliure com formant complexos amb diferents substrats i un inhibidor— amb una ressolució molt elevada (entre 1.9 i 2.2 Å; Mikami et al., 1993; Mikami et al., 1994) i també es coneixen un bon nombre de seqüències (tant d'ADN com de proteïna). A més, el nivell d'homologia entre les seves seqüències —entre el 99 i el 28%— és l'idoni per a fer un estudi evolutiu d'aquestes característiques.

La part inicial de la tesi —la que correspon a l'article Pujadas et al. (1996)— fa referència a l'anàlisi evolutiva de la β-amilasa. En aquest treball es desenvolupa un mètode que permet identificar de manera objectiva quins són els segments de la seqüència que presenten un grau menor d'evolució. Aquests segments són els responsables del manteniment de l'estructura terciària i de l'activitat enzimàtica. A més d'un anàlisi a nivell de seqüència, l'article també analitza quines repercusions ha tingut l'evolució en l'estructura. L'anàlisi estructural de la β-amilasa ha permés identificar un canvi conformacional —associat a la formació del complex enzim-substrat o enzim-inhibidor— que afecta a la tira β_6. El canvi conformacional involucra al segment ^{341}FTC^{343} i du associat un canvi de

rotàmers per la Thr_{342} i la Cys_{343}. L'anàlisi de la transició conformacional constitueix la segona part de la tesi i correspon a l'article Pujadas i Palau (1997). El darrer capítol de la tesi descriu les biblioteques de rotàmers depenents de l'estructura secundària, la seva validació i la metodologia emprada a l'hora d'obtenir-les. Els canvis de rotàmer que afecten a la Thr_{342} i la Cys_{343} s'han avaluat amb la biblioteca corresponent a les tires β. El text d'aquest capítol també està en anglès perquè properament formarà part d'una pàgina Web especialitzada en rotàmers (http:argo.urv.es/~pujadas).

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8:
RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

Resultats i discussió.

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8:
RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

Evolució de les β-amilases.

PROTEINS: Structure, Function, and Genetics 25:456–472 (1996)

Evolution of β-Amylase: Patterns of Variation and Conservation in Subfamily Sequences in Relation to Parsimony Mechanisms

Gerard Pujadas, Flora M. Ramírez, Ricard Valero, and Jaume Palau
Unitat de Biotecnologia Computacional, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Tarragona 43005, Catalonia, Spain

ABSTRACT **Soybean and sweet potato β-amylases are structured as α/β barrels and the same kind of folding may account for all known β-amylases. We provide a comprehensive analysis of both protein and DNA (coding region) sequences of β-amylases. The aim of the study is to contribute to the knowledge of the evolutionary molecular relationships among all known β-amylases. Our approach combines the identification of the putative eightfold structural core formed by β-strands with a complete multialignment analysis of all known sequences. Comparing putative β-amylase $(\alpha/\beta)_8$ cores from plants and microorganisms, two differentiated versions of residues at the packing sites, and a unique set of eight identical residues at the C-terminal catalytical site are observed, indicating early evolutionary divergence and absence of localized three-dimensional evolution, respectively. A new analytical approach has been developed in order to work out conserved motifs for β-amylases, mostly related with the enzyme activity. This approach appears useful as a new routine to find sets of motifs (each set being known as a fingerprint) in protein families. We demonstrate that the evolutionary mechanism for β-amylases is a combination of parsimonious divergence at three distinguishable rates in relation to the functional signatures, the barrel scaffold, and α-helix-containing loops.**

Key words: protein database, DNA database, PROSITE, fingerprints, PRINTS, α/β-barrel, protein evolution

INTRODUCTION

β-Amylase (EC 3.2.1.2) (1,4-α-D-glucan maltohydrolase) is an exo-acting carbohydrolase which hydrolyzes the (1,4)-α-D-glucosidic linkages of starch, glycogen, and maltooligosaccharides.[1] The enzyme liberates stepwise maltose units with β-anomeric configuration from the non-reducing end of the polysaccharide chain, producing β-maltose and β-limit dextrin as the final products. It is found in various higher plants[2,3] and in some gram-positive spore-forming bacteria.[4] This enzyme has a high biotechnological interest.[5,6]

At present, several cDNA, genomic DNA and protein β-amylase sequences are known from the following biological species: a) in plants: *Arabidopsis thaliana* (mouse-ear cress),[7] *Glycine max* (soybean),[8–11] *Hordeum vulgare* (barley),[12–14] *Ipomoea batatas* (sweet potato),[15–17] *Oryza sativa* (rice) (Chen, unpublished data), *Secale cereale* (rye),[18,19] *Zea mays* (maize)[20]; b) in microorganisms: *Bacillus cereus*,[21] *Bacillus circulans*,[22] *Bacillus polymyxa*,[23–25] and *Clostridium thermosulfurogenes*.[26] A number of studies have shown that β-amylase of mature kernels is found in multiple isoforms.[10,13,17,18,19,23,25,27–45]

Structures of the *Glycine max* β-amylase, either free or complexed with substrate or inhibitor have been recently resolved at a high quality resolution (from 1.9 Å to 2.2 Å)[46,47] and deposited in the Brookhaven Protein Data Bank (BPDB)[48] coded as 1BYA, 1BYB, 1BYC, 1BYD, and 1BTC. According to Mikami et al.,[46] the protein consists of a core with an $(\alpha/\beta)_8$ supersecondary structure, plus a smaller globular region formed by long loops (L3, L4, and L5) extending from β-strands β 3, β 4, and β 5. Between the two regions there is a cleft that opens into a pocket whose floor contains the postulated catalytic center near the carboxyl group of Glu 186.[47,49–51] Recently, the central role of loop L3 in catalysis has been demonstrated.[47] A difference in the conformation of a segment of loop 3 (residues 96–103) distinguishes unliganded soybean β-amylase (SBA) from SBA/maltose-H, SBA/maltose-L, and SBA/maltal, which have saccharides bound at the active site (this conformational change is not observed when the inhibitor α-cyclodextrin binds to the enzyme). When saccharides bind to the active site, loop 3 adopts a closed conformation and forms part of a structural cover over parts of the ligands

Received August 15, 1995; revision accepted February 8, 1996.
Address reprint requests to Jaume Palau, Unitat de Biotecnologia Computacional, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Plaça de la Imperial Tàrraco, 1, Tarragona 43005, Catalonia, Spain.

and reaction center that otherwise would be exposed to solvent.

More recently, the three-dimensional structure of the tetrameric β-amylase from sweet potato has been determined by molecular replacement methods using the monomeric structure of soybean as the starting model (with a current R-value of 20.3% and a resolution range of 8–2.3 Å).[52]

In this article, we present a computational study that provides insight into the evolutionary relationship among the different β-amylases (19 sequences, either from different biological sources, or from isoforms of a given species). The analysis is based on the sequence segments that comprise the $(\alpha/\beta)_8$ cores and the intercalated loops (as defined by the three-dimensional structure worked out for *Glycine max* β-amylase). We develop a multi-alignment strategy that considers protein and DNA sequences in parallel and we combine this approach with a topological study of the layered residues present in the putative $(\alpha/\beta)_8$ cores for the different molecular species of the whole protein sample. In addition to the results for β-amylase reported in this article, our work sheds some light on the evolutionary inconsistencies reported for proteins with an α/β barrel three-dimensional structure.[53–56]

MATERIALS AND METHODS

Protein and DNA sequences and three-dimensional data were imported from on-line releases available in current databases, using the Netscape navigator application v1.12N installed in a Macintosh personal computer and connected to WWW servers via the Internet network. All available β-amylase sequences were taken from the following databases: Swiss-Prot[57] (http://expasy.hcuge.ch/sprot/sprot-top.html), PIR[58] (http://www.gdb.org/Dan/proteins/pir.html), NRL-3D[59] (http://www.gdb.org/Dan/proteins/nrl3d.html), EMBL Nucleotide Data Base[60] (http://www.ebi.ac.uk/queries/queries.html), and GenBank[61] (http://www2.ncbi.nlm.nih.gov/genbank/query_form.html). Protein sequences and translated ORF from DNA sequences were tested for 100% identity and redundant entries or sequence fragments (not covering the entire segment object of our study) were removed from the sample. As a result of the selection, a working database was set up, consisting of 19 non-identical β-amylase protein sequences which correspond to 18 β-amylase translatable-ORF nucleic acid sequences (up to now, the DNA sequence for the AMYB_IPOBA β-amylase variant has not been worked out). We validated our amino acid sequence database subset by comparison with a β-amylase collection of sequences (17 entries) imported from the on-line OWL database[62] of non-redundant protein sequences (worked out from a server installed at the Johns Hopkins University, Baltimore, MD) (http://www.gdb.org/Dan/proteins/owl.html). Only minor differences were found, but they were relevant enough to be considered for our purposes: BLYBA (from *Hordeum vulgare*), S74932 (from *Ipomoea batatas*), and SOYBETAA (from *Glycine max*) were taken as differentiated isoforms with respect to other variants of the same species; AMYB_SECCE present in the OWL databank was not taken by us since it is a fragment sequence that should be excluded under the criteria chosen in this article. The GenBank code was preferred for labeling the selected entries (exceptions had to be made for S74932 from EMBL databank and AMYB_IPOBA from Swiss-Prot databank, since equivalent entries were not found within GenBank). Taxonomical criteria were used to group the different molecular species: a) prokaryota, bacillaceae: *Bacillus cereus, Bacillus circulans, Bacillus polymyxa, Clostridium thermosulfurogenes;* b) eukariota, planta: *Arabidopsis thaliana, Glycine max, Ipomoea batatas, Hordeum vulgare, Oryza sativa, Secale cereale,* and *Zea mays*. If any, isoforms from the same source (*Bacillus polymyxa, Glycine max, Ipomoea batatas, Hordeum vulgare,* and *Oryza sativa*) were also considered. The structural information needed was taken from X-ray diffraction data[46,47] for soybean β-amylase (BPDB entries: 1BYA, 1BYB, 1BYC, 1BYD, and 1BTC; resolution: 2.2 Å, 1.9 Å, 2.2 Å, 2.2 Å, and 2.0 Å, respectively; R-factor: 0.169, 0.149, 0.149, 0.145, 0.177, respectively). Sequences corresponding to 1BYA, 1BYB, 1BYC, 1BYD, 1BTC, and GMAMYB are identical.

Multiple sequence alignments were carried out for each subset (protein and DNA sequences) of the working database by applying the Clustal algorithm[63] and using the commercial program MEGALIGN v 3.02 from the Lasergene software package (1995, DNASTAR, Inc., London, UK) running in a Power Macintosh 6100/60. All sequences to be aligned were previously shortened giving rise to informationally homologous segments (formed by amino acid residues or nucleotides) corresponding to the complete α/β barrel and the internal loops. The amino acid segment of reference was that for GMAMYB (from residue 13 to residue 425); the starting and ending points for all other segments of the database were easily obtained by local alignments against GMAMYB. Joined "sequences" of variable lengths (54 residues for β-strands, 74 residues for conserved motifs, and the remaining residues for the α-helix-containing loops) were built from the corresponding multi-aligned distal segments. Phylogenetic trees and SI (similarity index) from these "sequences" were worked out by applying available MEGALIGN subroutines, that calculate the SI parameter between two sequences (i and j) based on the method of Wilbur and Lipman[64] with values for a gap penalty, K-tuple, number of top diagonals and window size of 3, 1, 5, and 5, respectively. The SI is calculated as the number of exactly

matching residues in this alignment minus a "gap penalty" for every gap that was introduced. The result then is expressed as a percentage of the shorter sequence length. Multiple alignment parameters (fixed and floating gap penalties) had both a value of 10. The protein weight matrix was PAM 250.

Consensus sequences in a PROSITE syntax[65] from a multi-alignment of the whole set of β-amylase molecular species were worked out following a novel procedure that takes into account the amino acid alphabets proposed by Karlin et al.[66] and the following criteria: a) the minimal length of a segment is six residues; b) for the whole sample, any position within a given segment must display a coherent structural character (either internal and/or ambivalent, or external and/or ambivalent); c) if b) is fulfilled, no more than two characters of the eight-letter chemical alphabet should be allowed for each position in the whole sample.

The previously described consensus sequences have been searched in the on-line releases of PROSITE[65] (http://expasy.hcuge.ch/sprot/prosite.html) and PRINTS databases[67,68] (http://www.biochem.ucl.ac.uk/~attwood/PRINTS/PRINTS.html).

Operating in a workstation HP Apollo 9000/400 PVRX, sets of four CA atoms of the respective alternate barrel-forming β-strands, located in roughly the same plane perpendicular to the barrel axis, were identified for SBA (1BYA structure). Two independent procedures were used that imply visualization and distance calculations, respectively. The commercial program used for visual identification was KGNGRAF (Motecc-90 package, IBM Corporation). Only backbone and side chain atoms from alternate strands, with CA atoms facing the interior of the barrel, were considered in order to visualize the different planes. Rotation of alternate strands around the axis of the barrel helped in the characterization of residues forming the different parallel CA-layers (hereafter named layers). As a confirmatory procedure, calculations using distances between CA atoms from residues at the interior of the barrel were performed using the program DISTCALC (written in FORTRAN 77), on the following basis: a) the four CA atoms forming a layer must be the closest ones within the whole set of residues; b) each residue from alternate strands participates only in one layer; and c) all layers must be displayed in a parallel distribution, being perpendicular to the barrel axis. By minimizing the summation values of CA-CA distances from all four-residue combinations, the layered residues were worked out in a precise manner. Both procedures have been positively tested for the α/β barrel proteins used by Lesk et al.[69]

Hydropathy values attributable to layered residues within the barrel were determined, using the Kyte and Doolittle hydropathic scale,[70] for all β-amylase molecular species under the assumption that residues multi-aligned with the eight β-strands from the crystallized SBA,[46,47] form in all cases the hyperboloid characteristic of an α/β barrel protein. The summation for each layer of the four hydropathy values resulted in global values measuring the hydropathy through the barrel axis. These values were plotted along the layer level of the hyperboloid.

RESULTS

Protein and DNA Phylogenetic Analysis Distinguishes, Within the Two Subgroups of Plant and Microbial β-Amylases, a Number of Protein Motifs

Nineteen protein-sequence segments of β-amylase were multi-aligned and the results were compiled as shown in Figure 1 and in Table I. Fig. 1 displays a multialignment of all β-amylase segments comprising residues that align with the segment from residue 13 to residue 425 of the crystallized β-amylase (GMAMYB) from *Glycine max*.[46,47]

Taking advantage of the structural and chemical alphabets reported by Karlin et al.,[66] and following criteria described in Materials and Methods, we describe a concept of structurally similar segments which define a set of conserved protein motifs presents in all β-amylase sequences of the different species. A motif is defined as an almost invariable consensus sequence (being representative of the whole set of segments aligned at a given site of a protein sequence multi-alignment) written in PROSITE syntax,[65] which is formed by six or more consecutive residues accomplishing the two following rules: a) Using the structural alphabet described in the review of Karlin et al.[66] (external: R, K, H, D, E, N, Q; internal: L, V, I, F, M, and ambivalent: A, C, G, P, S, T, W, Y), drastic structural changes (i.e., substitutions of an internal residue by an external one, or vice versa) are forbidden for the extension of the motif. b) If a) is accomplished, no more than two changes, throughout the eight-letter chemical alphabet (acidic: D, E; aliphatic: A, G, I, L, V; amide: N, Q; aromatic: F, W, Y; basic: K, R, H; hydroxyl: S, T; imino: P; sulfur C, M, also described in Karlin et al.,[66] are admitted. The ensemble of protein motifs for a given protein (in our case β-amylase) forms a fingerprint, a definition borrowed from PRINTS database.[67,68] Only eight subsequences showing high local multiple similarity (accounting for 74 residues, which correspond to approximately 18% of the whole multi-aligned segment) were found. They are indicated in the first and the last rows of Figure 1. The fact that the eight segments bundle themselves at the C-terminal end of the β-barrel scaffold (where the core of enzyme activity is located) is remarkable (see Table I and Fig. 2).

Table I shows the eight motifs in a PROSITE syntax[65] for the whole set of proteins and for the two evolutionary groups: microorganisms and plants. Figure 2 shows their three-dimensional arrange-

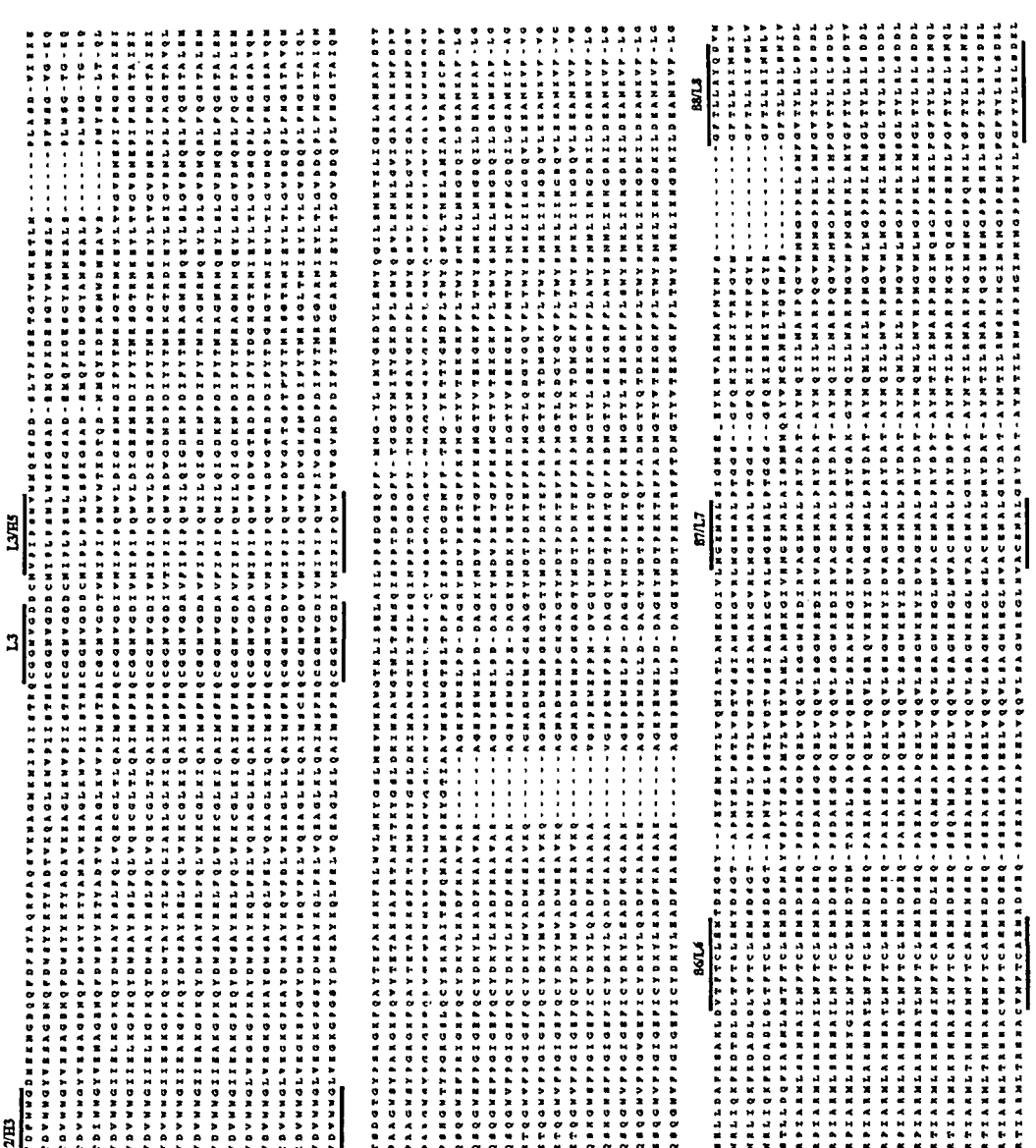


Fig. 1. Sequence multialignment of all known β-amylases (on-line databanks from July 29, 1995). Line marks denote high conservative sequence segments. The consensus sequences for the conserved segments are displayed in Table I.

TABLE I. Consensus Sequences for β-Amylases in a PROSITE Syntax (Specific Consensus for Different Evolutionary Groups Is Shown)*

Location	Whole sample		Microorganisms		Plants	
In β/L interzones						
β4/L4	[DK]-I-[EY]-[ILV]-[GS]-[CGL]-G-[AP]-[AS]-G-E-[LM]-R-Y-P-S-Y	0	K-I-Y-[IL]-S-G-G-P-[AS]-G-E-L-R-Y-P-S-Y	0	D-I-E-V-G-[CL]-G-[AP]-A-G-E-[LM]-R-Y-P-S-Y	0
β5/L5	[IL]-[AG]-[AI]-K-[IV]-[AS]-G-[ILV]-H-W	0	I-G-A-K-I-[AS]-G-[ILV]-H-W	0	[IL]-A-[AI]-K-[IV]-S-G-I-H-W	0
β6/L6	[ILMV]-[NT]-F-T-[AC]-[AL]-E-M	0	[LMV]-T-F-T-[AC]-L-E-M	0	[ILMV]-N-F-T-C-[AL]-E-M	0
β7/L7	[AN]-[CG]-E-N-A-L	20	N-G-E-N-A-L	12	A-[CG]-E-N-A-L	8
β8/L8	G-[FLV]-T-[LY]-L-R-[ILVY]-[NQS]-[DN]	1	G-F-T-L-L-R-[ILY]-[NQS]-[DN]	0	G-[FLV]-T-Y-L-R-[LV]-S-[DN]	0
In L/β/H interzones						
L'1/β2/H3	[DY]-[AG]-[IV]-[MT]-[ITV]-D-[FIV]-W-W-G	0	Y-[AG]-I-T-[TV]-D-[FIV]-W-W-G	0	D-G-[IV]-M-[IV]-D-V-W-W-G	0
In L regions						
L3	C-G-G-N-V-G-D	0	C-G-G-N-V-G-D	0	C-G-G-N-V-G-D	0
In L/H interzones						
L3/H5	[IV]-P-[IL]-P-[QS]-W-[ILV]	0	[IV]-P-[IL]-P-S-W-[LV]	0	I-P-I-P-Q-W-[IV]	0

*The values to the right of the signatures, generally 0, denote the number of sequences found for non-β-amylase entries at the merged LASERGENE protein databank as indicated in the Results section. Residues shown in boldface have been described to be relevant for the structure maintenance, active site, or substrate/inhibitor binding site (see also Table III).

ment. Several features become apparent: a) most of the subsequences start at the end of β strands and continue through the following loop (β4/L4, β5/L5, β6/L6, β7/L7, β8/L8); b) the loops L'1 and L3 are involved in the other three subsequences (L'1/β2/H3, L3, and L3/H5); c) with the exception of L3/H5, all other subsequences are involved in structural maintenance, active site, or substrate/inhibitor binding site; d) the two reported PROSITE signatures for β-amylases (PS00506, PS00679 in the PROSITE Databank release 13.0, the on-line current release), as defined by Bairoch,[65] are related, although not fully coincident, with L3 and β4/L4; e) with the exception of the short subsequence β7/L7, none shows full identity with a merged database (178.250 entries in May-1995 release, DNASTAR, Inc., London, UK), based on Swiss-Prot (release 31), NBRF-PIR, (release 44) and GenBank translated database (release 88); f) in the case of the β8/L8 signature, a nine-residue segment of an unrelated *Saccharomyces cerevisiae* hypothetical protein fortuitously fulfils the global signature but not the stringent ones for β-amylases from plants or from microorganisms; g) there is no β-amylase entry in the current on-line PRINTS database; h) Figure 2 reveals an L3 conformational change when a given substrate is located at the active site in comparison with other L3 conformations (for the free enzyme or for the enzyme-inhibitor complex).

From all the above considerations, it becomes apparent that the procedure worked out by us in order to identify structural similarities reveals, in the case of β-amylases, new protein motif subsequences that are linked to structural and/or functional regions of the molecule which forms a fingerprint. We suggest that our procedure could be applied as a general method to identify fingerprints to be incorporated in databases like PRINTS. The advantage of our procedure is that provides an objective way to define the variability and extension of each protein motif. In this sense, we believe that the new procedure will afford new insights with respect to observed structural signatures within the exponentially-growing collection of sequenced proteins and DNA-sequence-derived proteins (either recognizable or hypothetical proteins). We also remark that in the future, as more and more protein sequences will be added to the databases, a taxonomic fingerprint database, following similar patterns to those displayed in Table I, might be of interest for structural protein science.

In a next step of our work, we attempted to reconstruct evolutionary trees by using DNA and protein sequences from β-amylases. Following a unified way to understand evolution at two levels (natural selection of molecular phenotypic characters and random drift of selectively neutral mutants),[72] we analyzed the tree topology taking into account that in β-amylase sequences substitution rates (SR) vary over nucleotide sites. According to the SR, we classified

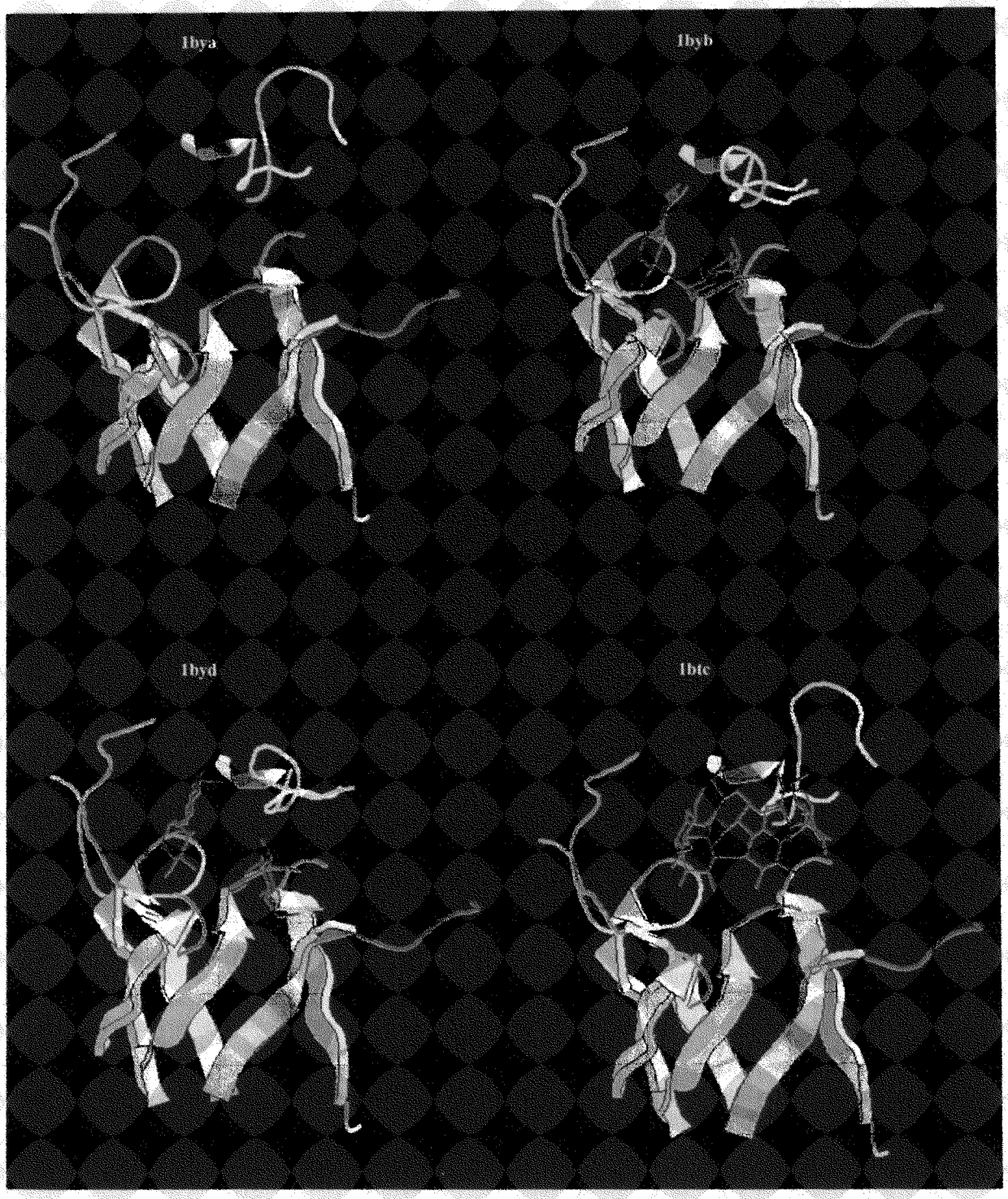


Fig. 2. The pictures show the eight β-amylase motifs (forming a fingerprint) bundling themselves under different conditions. The fingerprint is located at the C-terminal end of the β-barrel scaffold forming a core for the catalytical site. A significant change in the conformation of the L3 conserved motif is observed when the substrate binds to the enzyme (1byb and 1byd structures). The binding of an inhibitor does not change the L3 motif conformation (1btc structure). The BPDB code has been used to label the different structures. The pictures have been generated using the program MolView 1.1.[71]

sites into three classes: a) each of the eight motifs forming a fingerprint (low SR); b) each of the eight β-strands forming the barrel (intermediate SR); and the remainder of the sequence sites (high SR). Sites belonging to classes a) and b) are too short in sequence length. Therefore, we decided to join all sub-

sequences belonging to each class, in order to avoid sampling errors when considering local sites.[73] To this purpose, we made junctions of distal subsequences (giving rise to 74 joined residues for the fingerprint, 54 for the eight barrel-forming β-strands, and 291-318 for the rest of the molecule).

We performed four multi-alignments using the whole sequence (in A) and the three different joined sequences (in B, C, and D). In these alignments, one β-amylase sequence (encoded as SOYAMYLASE) was not considered since a short segment of it (QQESSDT, in β4/L4 of Fig. 1) shows a strong discrepancy with respect to all other local subsequences placed at the homologous conserved site (see comments below concerning possible frameshift errors). We obtained four phylogenetic trees (Fig. 3), as well as matrices of similarity indexes (SI). Instead of showing the matrices in full, we present the intervals from minimal to maximal SI values when comparisons are made among microorganisms, among plants and between the whole two sets of microorganisms and plants (Table II).

As expected, β-amylases are distributed in two different evolutionary subgroups (those from plants and those from bacteria). Figure 3A displays a clear-cut phylogenetic tree with a maximal evolutionary distance (in arbitrary units) of 37.4. The maximal value for evolutionary distances decreases to 16.8 and 27.5, in B and C respectively. Table II shows, for the complete sequence (type A), the SI ranges from 55 to 98% among bacteria β-amylases and from 64 to 99% among plant β-amylases. Comparisons between sequences from microorganisms and sequences from plants yield SI values ranging from 28 to 36%. In general, high overall SI values denote comparisons among isoforms of the same species or among β-amylases from taxonomically related species. Despite the low SI values in some cases, when sequences of one evolutionary subgroup are compared with sequences of the other subgroup, the nucleotide sequence pattern is recognizable for multi-alignment purposes.

The particular behavior of the four multi-alignments, as described and discussed above, indicates that the molecular evolution of β-amylases follows a parsimonious pattern of divergence, displaying a gradual evolutionary permissivity for distantly placed subsequences. The α-helix-containing loops shows the greatest permissivity, followed by the β-barrel scaffold (accounting for a structural level of evolutionary conservation) and, finally, by protein motifs related with the enzyme activity (accounting for a functional level of evolutionary conservation). Multi-alignment of the 18 sequence segments of β-amylase at the DNA level (the DNA sequence corresponding to AMYB__IPOBA being unknown) (results not shown) corroborates the divergent evolution along the whole sequence.

TABLE II. Ranges of Similarity Index (SI) Among β-Amylases Derived From Multialignments of: A) the Whole Sequence; B) the Joined Signatures; C) the Joined β-Strands; and D) the Joined α-Helix-Containing Loops

Sequence type	Among microorganisms	Among plants	Between microrganisms and plants
A	55–98	64–99	28–36
B	82–99	84–100	61–70
C	74–100	70–100	39–50
D	47–98	59–99	20–28

The Combined DNA and Protein Multi-Alignments Reveal Discrepancies Among Short Subsequences of β-Amylases

As pointed out above, a number of DNA sequence segments of β-amylase (18 sequences) were multi-aligned and the results were compiled in a similar way to those of protein sequences (results not shown). Combined DNA and protein alignments facilitated local protein matching. Multiple isoforms of β-amylase have been found from different sources by several authors (see introduction references). In general, sequence differences among isoforms, at the protein level, are merely due to non-silent nucleotide changes among codons that originate amino acid residue changes (e.g., RICAMYBA and RYCAMYBB isoforms from *Oryza sativa*).

Careful observation of the DNA results indicates discrepancies in subsequences which, in principle, could be nothing but sequencing artifacts accounting for the reported 1% sequence error rate within the primary structure databases.[74,75]

For example, the subsequence QQESSDT in β4/L4 of SOYAMYLASE (PAGELRY in β4/L4 of both GMAMYB and SOYBETAA, the other known soybean β-amylases) breaks a reported consensus pattern (PROSITE release 13.0 and updates) related with the enzyme function (entry code: PS00679: G-x-[AS]-G-E-[ILMV]-R-Y-P-S-Y, the Glu_{186} residue being characteristically involved in catalysis) (Fig. 4A). The DNA multi-alignment corresponding to this short segment of SOYAMYLASE sequence may suggest that the variation is due to a DNA-sequencing artifact (not detected by the authors and, possibly, related with the presence of repeated cytosines at both sides of the "odd-translatable" DNA short segment). Although less plausible, the presence of this unexpected short segment as a result of a double frameshift mutation (FSM) cannot be disregarded because, in this case, a Glu185 residue is present, offering a close alternative to the active Glu186.

In this systematic comparison of DNA and protein multi-alignments, we observed other local discrepancies among the whole set of sequences. None of them apparently affects any critical zone of the mol-

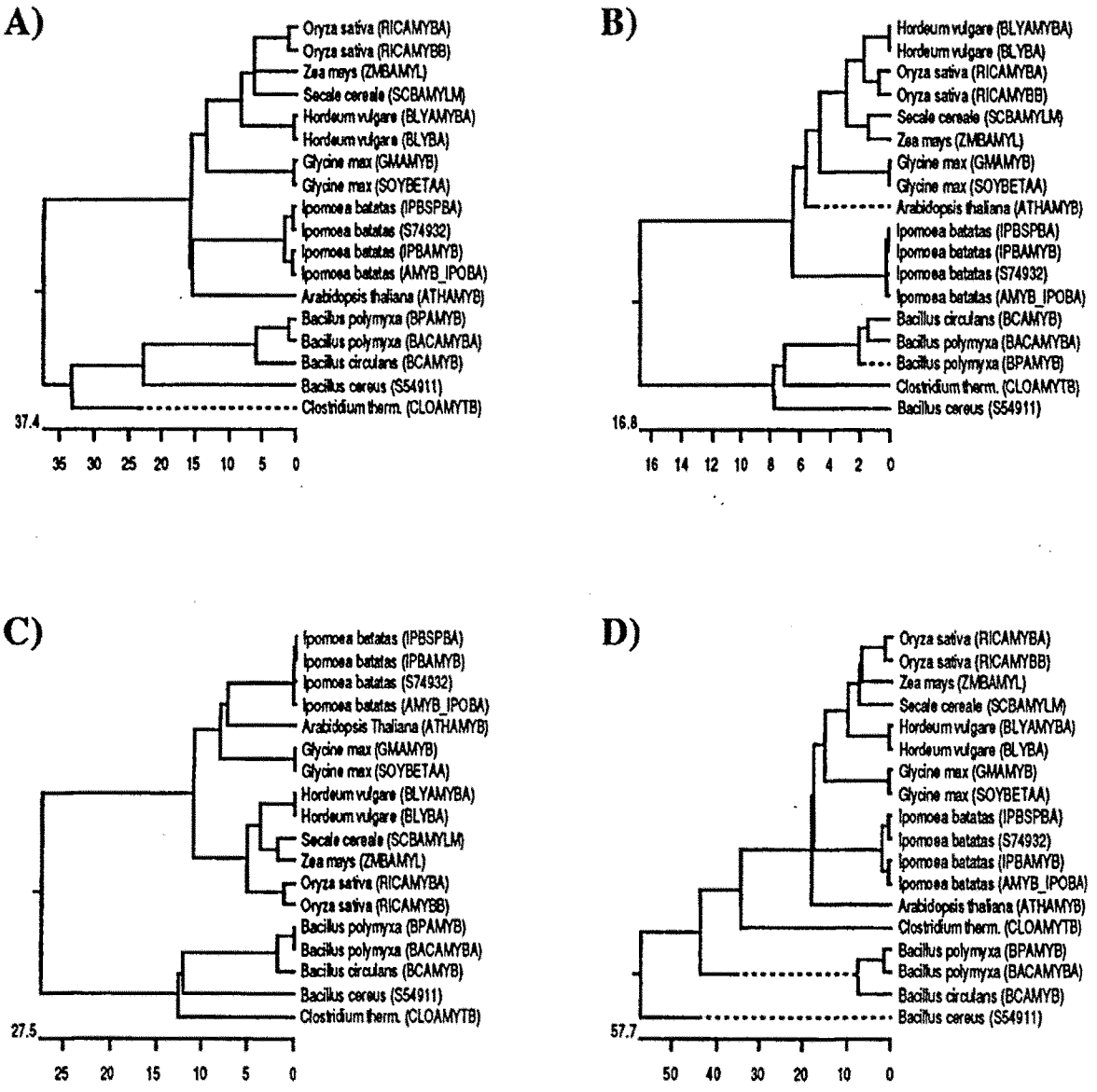


Fig. 3. Phylogenetic trees of β-amylases. **A:** For complete sequences as defined in Figure 1. **B:** For joined conserved motifs which constitute the β-amylase fingerprint. **C:** For joined β-strands. **D:** For joined α-helix-containing loops. Methods used to define segments in B, C, and D are described within the text.

ecule. For example, variations are observable for *Ipomoea batatas* β-amylase isoforms covering the amino acid residues from 255 to 263 located at a non-conservative region of L4/H9 and it yields the same amino acid subsequence in IPBSPBA and S74932 (NGTLQDGYGQVFL), and in IPBAMYB and AMYB_IPOBA (NGTYKTDMGKFFL) (Fig. 4B). Recently, Cheong et al.[52] attempt to resolve the observed discrepancy by elucidating the X-ray local structure without success in distinguishing by the density. In this case, the possibility for the existence of a compensated double frameshifting should also be considered.

Two other discrepancies in long sequence segments observed by us, none of them affecting any critical zone of the molecule, were the following: a) NKIFAGLKV for *Arabidopsis thaliana* (ATHAMYB) in comparison with all the other sequences from plants (e.g., NKAFLGCKV, for GMAMYB) (Fig. 4C); and b) KLISELAILPP for *Bacillus cereus* (S54911) as compared with other sequences from microorganisms (e.g., NLTSMSQISPP, for BCAMYB) (Fig. 4D). In these two cases, it is more plausible that the observed discrepancies reflect a compensated double FSM because for both sequence comparisons the DNA segments belong to

A

```
C T T G G C C - A G C A G G A G A G C T C A G A T A C C C C C T C T     (SOYAMYLASE)
C T T G G C C C A G C A G G A G A G C T C A G A T A - C C C C T C T     (GMAMYB)

181  Leu Gly | Gln Gln Glu Ser Ser Asp Thr | Pro Ser  191  (SOYAMYLASE)
     CTT GGC | CAG CAG GAG AGC TCA GAT ACC | CCC TCT
181  Leu Gly | Pro Ala Gly Glu Leu Arg Tyr | Pro Ser  191  (GMAMYB)
     CTT GGC | CCA GCA GGA GAG CTC AGA TAC | CCC TCT
```

B

```
A A C G G G A C C T T A C A A G A C G G A T A T G G G C A A G - T T T T C T T G     (IPBSPBA)
A A C G G G A C C T T A C A A G A C G G A T A T G G G C A A G - T T T T C T T G     (S74932)
A A C G G G A C - T T A C A A G A C G G A T A T G G G C A A G T T T T T C T T G     (IPBAMYB)

253  Asn Gly | Thr Leu Gln Asp Gly Tyr Gly Gln Val | Phe Leu  265  (IPBSPBA)
     AAC GGG | ACC TTA CAA GAC GGA TAT GGG CAA GTT | TTC TTG
253  Asn Gly | Thr Leu Gln Asp Gly Tyr Gly Gln Val | Phe Leu  265  (S74932)
     AAC GGG | ACC TTA CAA GAC GGA TAT GGG CAA GTT | TTC TTG
253  Asn Gly | Thr Tyr Lys Thr Asp Met Gly Lys Phe | Phe Leu  265  (IPBAMYB)
     AAC GGG | ACT TAC AAG ACG GAT ATG GGC AAG TTT | TTC TTG
```

C

```
A A C A A G A T C T T T G C T G G A C T - T A A A G T T     (ATHAMYB)
A A C A A A - G C T T T C C T G G G C T G T A A A G T C     (GMAMYB)

284  Asn Lys | Ile Phe Ala Gly Leu | Lys Val  292  (ATHAMYB)
     AAC AAG | ATC TTT GCT GGA CTT | AAA GTT
282  Asn Lys | Ala Phe Leu Gly Cys | Lys Val  290  (GMAMYB)
     AAC AAA | GCT TTC CTG GGC TGT | AAA GTC
```

D

```
A A A C T G A T A T C A G A G T T A G C A A T T T T A - - C C A C C A     (S54911)
A A C C T C - - A C C A G T A T G A G C C A G A T T A G T C C T C C A     (BCAMYB)

253  Lys Leu | Ile Ser Glu Leu Ala Ile Leu | Pro Pro  263  (S54911)
     AAA CTG | ATA TCA GAG TTA GCA ATT TTA | CCA CCA
252  Asn Leu | Thr Ser Met Ser Gln Ile Ser | Pro Pro  262  (BCAMYB)
     AAC CTC | ACC AGT ATG AGC CAG ATT AGT | CCT CCA
```

Fig. 4. Multialigned β-amylase subsequences displaying putative FSM. **A** and **B** refer to isozyme FSM. **C** and **D** FSM include sequences from different sources. FSM in β-amylase affects plants (A, B, and C) and microrganisms (D). Residues affected by an FSM are shown inside a box. The AMYB_IPOBA sequence is not included in B since its DNA sequence is unknown. The FSM shown in A would move the catalytical Glu one position upstream with respect to the crystallized GMAMYB sequence.

different sources, they are flanked by relatively well aligned regions and the presence of compensated gaps improves the local similarity between sequences.

Summarizing what is said in the above paragraphs, the observed discrepancies could be due merely to sequencing errors. However, some of them could be the result of a frameshift mutation, i.e., a one-nucleotide deletion/insertion mechanism causing an alteration of the reading frame (as a source of amino acid residue variations in the translated protein sequence), followed by the presence of a second deletion/insertion mutation nearby, allowing a recovery of the expected open reading frame (ORF) of

the nucleotide sequence. In any case, mutational frameshiftings, as an evolutionary mechanism in β-amylases, remains an open question which demands independent evidences other than sequence multi-alignment.

Topological and Multi-Alignment Studies Identify Two Differentiated Versions of Residues at the Packing Sites in the Interior of the α/β-Barrel Core

It has been described for several α/β barrel proteins that residues from the β-sheet that point towards the interior of the barrel core form a set of parallel layers (perpendicular to the barrel axis). Residue interactions in this layers are essential for structural maintenance and for the packing of the hyperboloid scaffold.[69,76,77] Each layer is formed by four residues that belong to alternate strands (one residue each strand). Up to now, studies on the packing of residues into the interior of the core have only been carried out for spinach glycolate oxidase (BPDB entry: 1GOX), *Rhodospirillum rubrum* ribulose-1,5-bisphosphate carboxylase-oxygenase (BPDB entry: 5RUB), and chicken triose phosphate isomerase (BPDB entry: 1TIM).[69]

We have now studied the layering of residues of the barrel in the case of soybean β-amylase (1BYA). Based on a procedure described in Materials and Methods that involves the shortest distances among sets of four CA atoms of the alternate barrel-forming β-strands, six layers can be considered for β-amylase structure (see Fig. 5A). Layers 3, 4, and 5 are well-packed allowing the formation of an $(\alpha/\beta)_8$ core. Beyond these layers there are regions in which four CA-atoms appear in roughly the same plane perpendicular to the barrel axis (layers 1, 2, and 6), but the attached residues are not packed together.

The more relevant clustered contacts, displayed between hydrophilic residues, are: Glu(OE2)178 and Asn(OD1)340 at 2.1Å; Glu(OE2)178 and Lys(CE) 295 at 2.8Å; Lys(NZ)295 and Glu(OE2)380 at 2.7Å; Asn(OD1)340 and Glu(OE2)380 at 3.0Å; and Thr(OG1)417 and Gln(NE2)87 at 3.1Å. Hydrophilic residues of the same packed layers (3, 4, and 5) do also interact with isolated water molecules at distances lesser than, or equal to, 3.0 Å. We found similar results (not shown) for the complexed β-amylase crystallized, indicating that the binding of substrates or inhibitors to the enzyme does not affect its internal packing.

The above results on packed layers are in agreement with results of Lesk et al.[69] which found three well-packed internal layers in different $(\alpha/\beta)_8$ barrel proteins. The other three layers (1, 2, and 6) may be merely considered as forming part of the overall topology of a large hyperboloid, in concomitance with the fact that the eight parallel β-strands display a length from 5 to 8 residues. We carried out a multi-alignment study for the amino acid residues from all β-strand subsequences integrated in the layered hyperboloid, as compared with the results obtained in 1BYA (GMAMYB in our code). Such an approach is similar to that used to obtain information about the maintenance of catalytic sites within β-amylase sequences,[10,15,17,21,26,46,51,78] but in our study we are dealing with the layered residue distribution. In the case of *Ipomoea batata*, our attribution of residues located at the interior of the hyperboloid (whether or not they participate in the packing) coincide with those indicated in a recent article.[52]

Results shown in Figure 5B for β-amylases from plants indicate that layer 1 is highly variable at the four positions (Gln, Ser, and Pro, in strand 1; Lys, Thr, in strand 3; Asn, Gln, Lys, in strand 5; Asn, Asp, His, Arg, Glu, in strand 7). Variations in layers 2 and 3 are restricted to one position (Thr, Ser, Ile, Cys, for strand6/layer2; Tyr, Asn, for strand1/layer3). Layers 4 to 6 are identical in all cases, thus indicating that the corresponding layered residues may be important for the stability of the β-barrel and/or for the enzymatic functional role, since a number of them have been reported to have non-carbon side-chain atoms located up to 3 Å from substrate.[47]

As Figure 5C shows, in the case of bacterial β-amylases, most residues from layers 1 to 3 undergo evolutionary variations. Residues, mainly from layers 3 and 4, diverge dramatically in comparison with plant β-amylases (this is the case for residues in strand5/layer3, strand7/layer3, strand2/layer4, and strand6/layer4 as highlighted in Figure 5B,C), suggesting that different modes of interaction have been selected for the packing sites of the hyperboloid scaffold for the two types of β-amylases. However, by comparing Figure 5B and C, it is clear that all residues from layers 5 and 6 are identical for any β-amylase from plants or microorganisms, which reinforces the view that the eight residues forming these two layers are essential for β-amylase structure and/or functional role. The mutational analysis of the eight codons for these residues reveals, within the whole set of DNA sequences, the presence of nearly all possible silent mutations giving rise, as a consequence, to only synonymous translations (results not shown).

Variations in Hydropathy Values Inside the Hyperboloid Suggest a Mechanism for the Stabilization and Reinforcement of the β-Amylase Scaffold

Following an idea described by Janecek[79] for α-amylases, we have developed a simple method to estimate variations of hydropathy through the barrel axis. The values were taken from a standard hydropathic scale as reported by Kyte and Doolittle,[70] and the procedure consists in adding the four resi-

A)

	Strand 1	Strand 2	Strand 3	Strand 4	Strand 5	Strand 6	Strand 7	Strand 8
Layer 1	Pro 13		Thr 85		Lys 291		Arg 376	
Layer 2		Gly 49		Asp 176		Ile 338		Gly 415
Layer 3	Tyr 15		Gln 87		Ala 293		Ala 378	
Layer 4		Met 51		Glu 178		Asn 340		Thr 417
Layer 5	Met 17		Ile 89		Lys 295 #		Glu 380 #	
Layer 6		Asp 53 #		Gly 180		Thr 342 #		Leu 419

B)

	Strand 1	Strand 2	Strand 3	Strand 4	Strand 5	Strand 6	Strand 7	Strand 8
Layer 1	Gln, Ser, Pro		Lys, Thr		Gln, Asn, Lys		Asp, Asn, His, Arg, Glu	
Layer 2		Gly		Asp *		Thr, Ser, Ile, Cys		Gly
Layer 3	Tyr, Asn		Gln		Ala *		Ala *	
Layer 4		Met *		Glu		Asn *		Thr
Layer 5	Met		Ile		Lys #		Glu #	
Layer 6		Asp #		Gly		Thr #		Leu

C)

	Strand 1	Strand 2	Strand 3	Strand 4	Strand 5	Strand 6	Strand 7	Strand 8
Layer 1	Gln, Lys		Lys		Arg, Pro		Arg, Val	
Layer 2		Ala, Gly		Ile *		Asp, Ala		Gly
Layer 3	Ser, Phe, Tyr		Val, Ile		Gly *		Asn *	
Layer 4		Thr *		Leu, Ile		Thr *		Thr
Layer 5	Met		Ile		Lys #		Glu #	
Layer 6		Asp #		Gly		Thr #		Leu

Fig. 5. CA-layered residues located inside the β-barrel. Layer 4 is situated at the central plane of the hyperboloid. **A:** Layered residues in the crystallized SBA. **B:** Putative layered residues in the sample plants. **C:** Putative layered residues in the sample microorganisms. *, Denotes residues conserved within an evolutionary group but not between groups. #, Denotes residues with non-carbon side-chain atoms located at distances less than 3Å from substrate in 1BYB. Glu 380 is involved in the catalytic action.[47]

due hydropathic values for each layer. The results are displayed in Figure 6 and the graphs represent the variation of hydropathy through the layer level for β-amylases from microorganisms (Fig. 6A) and from plants (Fig. 6B) from the N to the C-terminal side of the barrel. Hydrophilicity is very high in layer 1, in all cases. There is a rise in hydropathy in layer 2, even reaching positive values in the case of microorganisms. The clear-cut distinction between plants and microorganisms is represented by hydropathy in layer 4, which again reinforces the view that evolution has separated the two subsets of β-amylases. Of special interest is the evidence of higher hydrophobic value in layer 2 and 3 in the case of *Clostridium thermosulfurogenes*, in comparison with the genus *Bacillus*. This finding might be related to a higher stabilization of the hyperboloid scaffold in the case of the thermophile bacteria.

DISCUSSION

The Two Phylogenetic Subgroups of β-Amylases

The degree of homology of *B. cereus* β-amylase compared with other β-amylases from different microorganisms and plants has been studied elsewhere[21] and the results of these authors are consistent with our finding for the region that excludes the extra raw-starch-binding domain present in a wide variety of microorganisms.

Results displayed in Figures 1, 3, 5, and 6, and also in Tables I and II, clearly show that a divergent evolution took place that separated β-amylases from plant and microorganisms in two subgroups. At a more localized level, divergences highlighted in layers 3 and 4 (see Fig. 5B,C) reinforces the above inference. The molecules evolved by divergence at different rates (Fig. 3), following a parsimonious

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8: RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

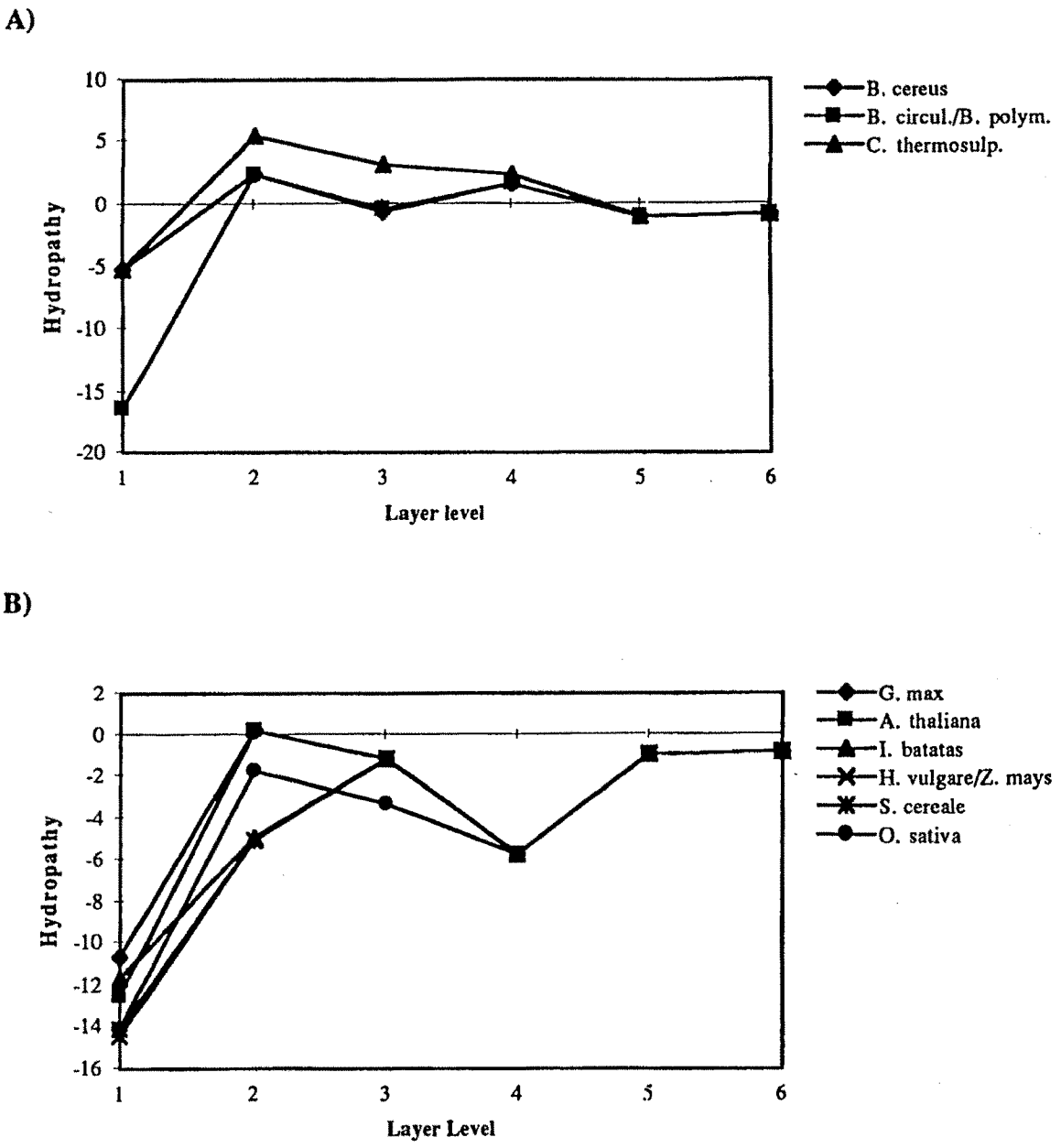


Fig. 6. Hydropathy plots through the layers of the barrel. **A:** For different microorganisms. **B:** For different plants.

mechanism of evolution. Such a parsimony pathway is corroborated by our results on β-amylase DNA multi-alignment (results not shown). The existence of eight conserved regions with stringent structural similarities (see Fig. 1 and Table I) that are related to functional and/or structural roles (see Table III and Fig. 2), and the identity of the eight residues in the critical layers 5 and 6 of the hyperboloid scaffold (Fig. 5), clearly point to a unique common ancestor for β-amylases.

The ensemble of phylogenetic branched trees for the three sets of joined subsequences (Fig. 3B,C,D) is rather well preserved (in branch distribution and in relative branch lengths) in comparison with the tree shown in Figure 3A. This overall similarity would favor the view that there is a consistency in the tree topology whatever be the nucleotide site considered.[73] Nonetheless, some observed discrepancies precludes us to establish, for β-amylases, relationships between evolutionary tree estimates at specific sites and true tree topologies.[73,81,82]

Clostridium thermosulfurogenes β-amylase may

ylases do not cross all taxonomic boundaries. Since protein sequence data grow at a very fast rate, we strongly recommend the incorporation of our two proposals for the sake of increasing the number of signatures.

CONCLUSION

The level of structural and/or functional constraints in a protein is thought to be determined by functionally outstanding sites as those critical for maintaining a given folding class, or involved in the active center of the molecule, or interacting directly with other molecules.[81] Our present results are completely consistent with such general ideas of protein evolution.

Farber[54] pointed out that the eight-stranded α/β-barrel (TIM barrel) is by far the most common shape for an enzyme, roughly one out of 10 enzymes having this domain. The title of Farber's article indicates that this structure is full of evolutionary inconsistencies and the author concludes that one cannot yet differentiate among the three different evolutionary pathways proposed for this group of proteins. Janecek,[55] by comparing the AMY superfamily with β-amylase reached no conclusion either, concerning the three possibilities (convergent evolution, divergent evolution, and exon combination). Considering Farber's and Janecek's statements, we decided in this article to study only one of such protein structures: the β-amylase enzyme, taking into account the great variety of sequences worked out up to now from many sources covering plants and microorganisms, and taking advantage of the fact that the structure has been elucidated in two cases.

According to the neutral theory of molecular evolution, the molecular evolutionary rate depends on mutation rate and on the degree of functional constraint against sequence variations.[72] Our results, showing different rates of parsimonious divergence, including some deletions and/or insertions, suggest that one should study the evolutionary pathway of the TIM barrel proteins one by one, before attempting to understand the evolution of the whole class of polypeptide folding.

The great interest in using α/β-barrel scaffold for biotechnological purposes also requires individualized studies of each molecular species in order to understand the structural basis of the mechanism by which apparently unrelated sequences undergo the same type of folding. Our routine revealing the presence of a bundle of conserved motifs (fingerprint) in β-amylases could become a useful tool that should be applicable to other systems. The observed conformational changes of the fingerprint, shown in Figure 2, results in a precise mechanism to adapt the substrate into the catalytical site. Whether such three-dimensional studies are extensible to a number of glycosyl hydrolases, also showing an $(\alpha/\beta)_8$ folding, is an issue that merits further attention.

ACKNOWLEDGMENTS

We thank Drs. Enric Querol and Thomas J. Smith for providing facilities and help in the use of the latest versions of the LASERGENE and MolView programs, respectively. We also thank the Unit of Physical-Chemistry of our University for facilities in using the Motecc-90 package. G.P. has been recipient of a fellowship (FI/93-129) from the Catalan Governmental Agency CIRIT (Generalitat de Catalunya).

REFERENCES

1. Robyt, J.F., Whelon, W.J. The β-amylase. In: "Starch and Its Derivatives." Radley J. (ed.). London: Chapman and Hall, 1968:430–476.
2. Bernfeld, P. Amylases, α and β. Methods Enzymol. 1:149–158, 1955.
3. Thoma, J.A., Spradlin, J.E., Dygert, S. Plant and animal amylases. In: "The Enzymes." 3rd edit., chap. 5. Boyer P.D. (ed.). New York: Academic Press, 1971:115–189.
4. Nanmori, T. Bacterial β-amylases (B. cereus, B. polymyxa, etc.). In: "Handbook of Amylases and Related Enzymes." The Amylase Research Society of Japan (ed.). Oxford: Pergamon Press, 1988:94–99.
5. Franks, F. Introduction. In: "Characterization of Proteins." Franks F. (ed.). Cambridge, UK: Pafra, Ltd., 1988: 1–8.
6. Fukazawa, C. Method for separation of beta-amylase. Patent number 05294341, Tsukuba, Japan, 1994.
7. Monroe, J.D., Salminen, M.D., Preiss, J. Nucleotide sequence of a cDNA clone encoding a beta-amylase from *Arabidopsis thaliana*. Plant Physiol. 97:1599–1601, 1991.
8. Mikami, B., Nomura, K., Morita, Y. N-terminal sequence of soybean beta-amylase. J. Biochem. 100:513–516, 1986.
9. Ehrlich, K.C., Montalbano, B.G. Nucleotide and deduced protein sequence of beta-amylase. Directly submitted to DNA Data Bank of Japan, 1992.
10. Totsuka, A., Fukazawa, C. Expression and mutation of soybean beta-amylase in *Escherichia coli*. Eur. J. Biochem. 214:787–794, 1993.
11. Adachi, M., Katsube, T., Mikami, B., Utsumi, S. Enzymatic and crystallographic analysis of soybean beta-amylase expressed in *E. coli*. Directly submitted to DNA Data Bank of Japan, 1995.
12. Kreis, M., Williamson, M., Buxton, B., Pywell, J., Hejgaard, J., Svendsen, I. Primary structure and differential expression of beta-amylase in normal and mutant barleys. Eur. J. Biochem. 169:517–525, 1987.
13. Yoshigi, N., Okada, Y., Sahara, H., Koshino, S. PCR cloning and sequencing of the beta-amylase cDNA from barley. J. Biochem. 115:47–51, 1994.
14. Yoshigi, N., Okada, Y., Sahara, H., Koshino, S. Expression in *Escherichia coli* of cDNA encoding barley beta-amylase and properties of recombinant beta-amylase. Biosci. Biotechnol. Biochem. 58:1080–1086, 1994.
15. Yoshida, N., Nakamura, K. Molecular cloning and expression in *Escherichia coli* of cDNA encoding the subunit of sweet potato beta-amylase. J. Biochem. 110:196–201, 1991.
16. Yoshida, N., Hayashi, K., Nakamura, K. A nuclear gene encoding beta-amylase of sweet potato. Gene 120:255–259, 1992.
17. Toda, H., Nitta, Y., Asanami, S., Kim, J.P., Sakiyama, F. Sweet potato beta-amylase. Primary structure and identification of the active-site glutamyl residue. Eur. J. Biochem. 216:25–38, 1993.
18. Rorat, T., Sadowski, J., Grellet, F., Daussant, J., Delseny, M. Characterization of cDNA clones for rye endosperm beta-amylase and analysis of beta-amylase deficiency in rye mutant lines. Theor. Appl. Genet. 83:257–263, 1991.
19. Sadowski, J., Rorat, T., Cooke, R., Delsensy, M. Nucleotide sequence of a cDNA clone encoding ubiquitous beta-amylase in rye (*Secale cereale L*). Plant Physiol. 102:315–316, 1993.
20. Wang, S., Lue, W., Chen, J. Sequence of a beta-amylase

cDNA from *Zea mays*. Directly submitted to the European Molecular Biology Laboratory Nucleotide Database/GenBank/DNA Data Bank of Japan databases, 1993.
21. Nanmori, T., Nagai, M., Shimizu, Y., Shinke, R., Mikami, B. Cloning of the beta-amylase gene from *Bacillus cereus* and characteristics of the primary structure of the enzyme. Appl. Environ. Microbiol. 59:623–627, 1993.
22. Siggens, K.W. Molecular cloning and characterization of the beta-amylase gene from *Bacillus circulans*. Mol. Microbiol. 1:86–91, 1987.
23. Kawazu, T., Nakanishi, Y., Uozumi, N., Sasaki, T., Yamagata, H., Tsukagoshi, N., Udaka, S. Cloning and nucleotide sequence of the gene coding for enzymatically active fragments of the *Bacillus polymyxa* beta-amylase. J. Bacteriol. 169:1564–1570, 1987.
24. Rhodes, C., Strasser, J., Friedberg, F. Sequence of an active fragment of *B. polymyxa* beta amylase. Nucleic Acids Res. 15:3934, 1987.
25. Uozumi, N., Sakurai, K., Sasaki, T., Takekawa, S., Yamagata, H., Tsukagoshi, N., Udaka, S. A single gene directs synthesis of a precursor protein with beta- and alpha-amylase activities in *Bacillus polymyxa*. J. Bacteriol. 171:375–382, 1989.
26. Kitamoto, N., Yamagata, H., Kato, T., Tsukagoshi, N., Udaka, S. Cloning and sequencing of the gene encoding thermophilic beta-amylase of *Clostridium* thermosulfurogenes. J. Bacteriol. 170:5848–5854, 1988.
27. Gertler, A., Birk, Y. Purification and characterization of a β-amylase from soybeans. Biochem. J. 95:621–627, 1965.
28. Tkachuk, R., Tipples, K.H. Wheat β-amylases. II. Characterization. Cereal Chem. 43:62–79, 1966.
29. LaBerge, D.E., Meredith, O.S. Chromatographic properties of barley and malt β-amylases. J. Inst. Brew. 75:19–25, 1969.
30. Visuri, K., Nummi, M. Purification and characterization of crystalline β-amylase from barley. Eur. J. Biochem. 28: 555–565, 1972.
31. Mikami, B., Aibara, S., Morita, Y. Distribution and properties of soybean β-amylase isozymes. Agric. Biol. Chem. 46:943–953, 1982.
32. Bureau, D., Daussant, J. Efficiency of a smooth description procedure for the purification of barley β-amylase using immunoaffinity chromatography. Biochimie 65:361–365, 1983.
33. Sadowski, J., Wiatroszak, I., Daussant, J. Variability of β-amylase isozymes within a collection of inbred lines or rye (*Secale cereale L*). Experientia 44:352–354, 1986.
34. Lundgard, R., Svensson, B. The four major forms of barley β-amylase. Purification, characterization and structural relationship. Carlsberg Res. Commun. 52:313–326, 1987.
35. Lin, T.P., Spilatro, S.R., Preiss, J. Subcellular location and characterization of amylases in *Arabidopsis* leaf. Plant Physiol. 86:251–259, 1988.
36. Mikami, B., Morita, Y. Soybean β-amylase. In: "Handbook of Amylases and Related Enzymes." The Amylase Research Society of Japan (ed.). Oxford: Pergamon Press, 1988:87–89.
37. Mikami, B., Morita, Y. Amino acid sequence and three dimensional structure of β-amylase. In: "Handbook of Amylases and Related Enzymes." The Amylase Research Society of Japan (ed.). Oxford: Pergamon Press, 1988:99–104.
38. Shewry, P.R., Parmar, S., Buxton, B., Gale, M.D., Liu, C.J., Hejgaard, J., Kreis, M. Multiple molecular forms of β-amylase in seeds and vegetative tissues of barley. Planta 176:127–134, 1988.
39. Daussant, J., Laurière, C. Detection and partial characterization of two antigenically distinct β-amylases in developing kernels of wheat. Planta 181:505–511, 1990.
40. Ji, E.S., Mikami, B., Kim, J.P., Morita, Y. Positions of substituted amino acids in soybean beta-amylase isoenzymes. Agric. Biol. Chem. 54:3065–3067, 1990.
41. Kohno, A., Shinke, R., Nanmori, T. Features of the beta-amylase isoform system in dry and germinating seeds of alfalfa (*Medicago sativa L*). Biochim. Biophys. Acta 1035: 325–330, 1990.
42. Daussant, J., Sadowski, J., Rorat, T., Mayer, C., Laurière, C. Independent regulatory aspects and post-translational modifications of two β-amylases of rye. Use of a mutant inbred line. Plant Physiol. 96:84–90, 1991.
43. Takekawa, S., Uozumi, N., Tsukagoshi, N., Udaka, S. Proteases involved in generation of beta- and alpha-amylases from a large amylase precursor in *Bacillus polymyxa*. J. Bacteriol. 173:6820–6825, 1991.
44. Daussant, J., Sadowski, J., Ziegler, P. Cereal beta-amylases: Diversity of the beta-amylase isozyme status within cereals. J. Plant Physiol. 143:585–590, 1994.
45. Ziegler, P., Daussant, J., Loos, K. Development of beta-amylase activity and polymorphism in wheat seedling shoot tissues. J. Exp. Bot. 45:1147–1155, 1994.
46. Mikami, B., Hehre, E.J., Sato, M., Katsube, Y., Hirose, M., Morita, Y., Sacchettini, J.C. The 2.0-Å resolution structure of soybean beta-amylase complexed with alpha-cyclodextrin. Biochemistry 32:6836–6845, 1993.
47. Mikami, B., Degano, M., Hehre, E.J., Sacchettini, J.C. Crystal structures of soybean beta-amylase reacted with beta-maltose and maltal: Active site components and their apparent roles in catalysis. Biochemistry 33:7779–7787, 1994.
48. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. The Protein Data Bank: A computer based archival file for macromolecular structures. J. Mol. Biol. 112:535–542, 1977.
49. Nitta, Y., Isoda, Y., Toda, H., Sakiyama, F. Identification of glutamic acid 186 affinity-labeled by 2,3-epoxypropyl alpha-D-glucopyranoside in soybean beta-amylase. J. Biochem. 105:573–576, 1989.
50. Uozumi, N., Matsuda, T., Tsukagoshi, N., Udaka, S. Structural and functional roles of cysteine residues of *Bacillus polymyxa* beta-amylase. Biochemistry 30:4594–4599, 1991.
51. Totsuka, A., Nong, V.H., Kadokawa, H., Kim, C.S., Itoh, Y., Fukazawa, C. Residues essential for catalytic activity of soybean beta-amylase. Eur. J. Biochem. 221:649–654, 1994.
52. Cheong, C.G., Eom, S.H., Chang, C., Shin, D.H., Song, H.K., Min, K., Moon, J.H., Kim, K.K., Hwang, K.Y., Suh, S.W. Crystallization, molecular replacement solution and refinement of tetrameric β-amylase from sweet potato. Proteins 21:105–117, 1995.
53. Farber, G.K., Petsko, G.A. The evolution of α/β barrel enzymes. Trends Biochem. Sci. 15:228–234, 1990.
54. Farber, G.K. An α/β-barrel full of evolutionary trouble. Curr. Opin. Struct. Biol. 3:409–412, 1993.
55. Janecek, S. Parallel β/α-barrels of α-amylase, cyclodextrin glycosyltransferase and oligo-1,6-glucosidase versus the barrel of β-amylase: Evolutionary distance is a reflexion of unrelated sequences. FEBS Lett. 353:119–123, 1994.
56. Reardon, D., Farber, G.K. The structure and evolution of α/β barrel proteins. FASEB J. 9:497–503, 1995.
57. Bairoch, A., Boeckmann, B. The Swiss-Prot protein sequence data bank: Current status. Nucleic Acids Res. 22: 3578–3580, 1994.
58. George, D.G., Barker, W.C., Mewes, H.W., Pfeiffer, F., Tsugita, A. The PIR-International protein sequence database. Nucleic Acids Res. 22:3569–3573, 1994.
59. Pattabiraman, N., Namboodiri, K., Lowrey, A., Gaber, B.P. NRL-3D: A sequence-structure database derived from the protein data bank (PDB) and searchable within the PIR environment. Protein Seq. Data Anal. 3:387–405, 1990.
60. Emmert, D.B., Stoehr, P.J., Stoesser, G., Cameron, G. The European Bioinformatics Institute (EBI) databases. Nucleic Acids Res. 22:3445–3449, 1994.
61. Benson, D.A., Boguski, M., Lipman, D.J., Ostell, J. GenBank. Nucleic Acids Res. 22:3441–3444, 1994.
62. Bleasby, A.J., Akrigg, D., Attwood, T.K. OWL—a non-redundant, composite protein sequence database. Nucleic Acids Res. 22:3574–3577, 1994.
63. Higgins, D.G., Sharp, P.M. Fast and sensitive multiple sequence alignments on a microcomputer. CABIOS 5:151–153, 1989.
64. Wilbur, W.J., Lipman, D.J. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. U.S.A. 80:726–730, 1983.
65. Bairoch, A. PROSITE: Recent developments. Nucleic Acids Res. 22:3578–3580, 1994.
66. Karlin, S., Bucher, P., Brendel, V., Altschul, S.F. Statistical methods and insights for protein and DNA sequences. Annu. Rev. Biophys. Biophys. Chem. 20:175–203, 1991.

67. Attwood, T.K., Beck, M.E., Bleasby, A.J., Parry-Smith, D.J. PRINTS—a database of protein motif fingerprints. Nucleic Acids Res. 22:3590–3596, 1994.
68. Attwood, T.K., Beck, M.E. PRINTS—a protein motif finger-print database. Protein Eng. 7:841–848, 1994.
69. Lesk, A.M., Brändén, C.I., Chothia, C. Structural principles of α/β barrel proteins: The packing in the interior of the sheet. Proteins 5:139–148, 1989.
70. Kyte, J., Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157:105–132, 1982.
71. Smith, T.J. MolView: A program for analyzing displaying atomic structures on the Macintosh personal computer. J. Mol. Graphics 13:122–125, 1995.
72. Kimura, M. "The Neutral Theory of Molecular Evolution." Cambridge: Cambridge University Press, 1983.
73. Yang, Z. Evaluation of several methods for estimating phylogenetic trees when substitution rates differ over nucleotide sites. J. Mol. Evol. 40:689–697, 1995.
74. Gish, W., States, D.J. Identification of protein coding regions by database similarity search. Nat. Genet. 3:266–272, 1993.
75. Claverie, J.M. Detecting frame shifts by amino acid sequence comparison. J. Mol. Biol. 234:1140–1157, 1993.
76. Chothia, C., Finkelstein, A.V. The classification and origins of protein folding patterns. Annu. Rev. Biochem. 59: 1007–1039, 1990.
77. Brändén, C.I., Tooze, J. Alpha/beta structures. In: "Introduction to Protein Structure." New York: Garland Publishing, 1991:43–57.
78. Mikami, B., Nomura, K., Morita, Y. Two sulfhydryl groups near the active site of soybean beta-amylase. Biosci. Biotechnol. Biochem. 58:126–132, 1994.
79. Janecek, S. Does the increased hydrophobicity of the interior and hydrophilicity of the exterior of an enzyme structure reflect its increased thermostability? Int. J. Biol. Macromol. 15:317–318, 1993.
80. Mikami, B., Sato, M., Shibata, T., Hirose, M., Aibara, S., Katsube, Y., Morita, Y. Three-dimensional structure of soybean beta-amylase determined at 3.0 Å resolution: Preliminary chain tracing of the complex with alpha-cyclodextrin. J. Biochem. 112:541–546, 1992.
81. Kuma, K., Iwabe, N., Miyata, T. Functional constraints against variations on molecules from the tissue level: Slowly evolving brain-specific genes demonstrated by protein kinase and immunoglobulin supergene families. Mol. Biol. Evol. 12:123–130, 1995.
82. Lake, J.A. Reconstructing evolutionary trees from DNA and protein sequences: Paralinear distances. Proc. Natl. Acad. Sci. U.S.A. 91:1455–1459, 1994.
83. Friedberg, F., Rhodes, C. Segments of amino acid sequence similarity in beta-amylases. Protein Seq. Data Anal. 1:499–501, 1988.
84. Uozumi, N. Functional roles of active site residues of *Bacillus polymyxa* beta-amylase. Ann. N.Y. Acad. Sci. 672: 24–28, 1992.

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8:
RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

Transicions conformacionals al centre actiu de les β-amilases.

Protein Science (1997), 6:2409–2417. Cambridge University Press. Printed in the USA.

Anatomy of a conformational transition of β-strand 6 in soybean β-amylase caused by substrate (or inhibitor) binding to the catalytical site

GERARD PUJADAS AND JAUME PALAU

Unitat de Biotecnologia Computacional, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Plaça de la Imperial Tàrraco, 1, Tarragona 43005, Catalonia, Spain

(RECEIVED March 3, 1997; ACCEPTED July 18, 1997)

Abstract

A computational study of the five soybean β-amylase X-ray structures reported so far revealed a peculiar conformational transition after substrate (or inhibitor) binding, which affects a segment of the β-strand 6 (residues 341–343) in the $(\beta/\alpha)_8$ molecular scaffold. Backbone distortions that involve considerable changes in the ϕ and ψ angles were observed, as well as two sharp rotamer transitions for the Thr_{342} and Cys_{343} side chains. These changes caused the outermost CA-layer (at the C-terminal side of the barrel), which is involved in the catalysis, to shrink. Our observations strongly suggest that the ^{341}FTC^{343} residue conformations in the free enzyme are not optimal for protein stability. Furthermore, as a result of conformational transitions in the ligand-binding process, there is a negative enthalpy change for these residues (−27 and −34 kcal/mol, after substrate or inhibitor binding, respectively). These findings support the proposed "stability-function" hypothesis for proteins that recognize a ligand (Shoichet BK, Baase WA, Kuroki R, Matthews BW. 1995. A relationship between protein stability and protein function. *Proc Natl Acad Sci USA 92*:452–456). They are also in good agreement with other experimental results in the literature that describe the role of the 341–343 segment in β-amylase activity. Site-directed mutagenesis focused on these residues could be useful for undertaking functional studies of β-amylase.

Keywords: backbone; beta-amylase; $(\beta/\alpha)_8$-barrel; conformational transition; glycosyl hydrolase; potential energy; rotamer library; side chain

β-Amylase (EC 3.2.1.2) is a glycosyl hydrolase that hydrolyzes the (1,4)-α-D-glucosidic linkages of a variety of substrates such as starch, glycogen, and maltooligosaccharides from the non-reducing end of their polysaccharide chain. The final products of the enzymatic action are β-maltose, with a β-anomeric configuration, and a β-limit dextrin. The enzyme is present in various higher plants (Thoma et al., 1971) and in some gram-positive spore-forming bacteria (Nanmori, 1988). In recent years, *Glycine max* enzyme structures, which are either free or form a complex with substrates or an inhibitor (Mikami et al., 1993, 1994), have been resolved at high-quality resolution and deposited in the Brookhaven Protein Data Bank (PDB) (Bernstein et al., 1977). The resulting structure was described as a core with a $(\beta/\alpha)_8$ supersecondary structure, and a smaller globular region formed by long loops (L3, L4, and L5) extending from β-strands β3, β4, and β5. A region of the L3 (residues 96–103) changes its conformation when a substrate binds to the enzyme and covers the reaction center (Mikami et al., 1994). This conformational change is not observed when α-cyclodextrin binds to the enzyme because the steric hindrance of the inhibitor molecule does not allow any conformational change of the L3 segment (Mikami et al., 1994; Pujadas et al., 1996).

In a recent paper (Pujadas et al., 1996), we reported that β-amylases can be characterized by a fingerprint formed by a set of eight distally conserved sequence motifs, which are grouped at the C-terminal end of the β-barrel scaffold (where the core of enzyme activity is located). Five of these motifs start at the C-ending tails of β-strands and continue through the next loop (β4/L4, β5/L5, β6/L6, β7/L7, β8/L8), whereas the other three are made up of L'1/β2/H3, L3, and L3/H5 subsequences. The total length of the fingerprint is 74 residues, which is about 15% of the soybean β-amylase (SBA) sequence. In the same paper, we also described the geometry of the $(\beta/\alpha)_8$-barrel as being formed by six CA-layers, each of which is defined by a set of four alpha carbon atoms (CA) of the respective alternate barrel-forming β-strands. These CA-atoms are located on roughly the same plane, perpendicular to the barrel axis. It should be pointed out that the eight residues that form layers 5 and 6 (which are in the C-terminal part of the barrel) are fully conserved throughout evolution both in plants and in microorganisms. It is also interesting to note that some of these

Reprint requests to: Jaume Palau, Unitat de Biotecnologia Computacional, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Plaça de la Imperial Tàrraco, 1, Tarragona 43005, Catalonia, Spain; email: palau@quimica.urv.es.

residues (Asp_{53}, Lys_{295}, Thr_{342}, and Glu_{380}) are involved in interactions with the substrate (Mikami et al., 1994).

In the present paper, we describe a clear-cut transition of the C-terminal part of β-strand six (β6 hereafter) when the ligand (either the substrate or the inhibitor) binds to the enzyme. We would like to emphasize the importance of this finding, because it shows three residues (Phe_{341}, Thr_{342}, and Cys_{343}) which appear to be involved, at least from a structural point of view, in the attachment of substrates or inhibitors to the catalytical site.

Results

Conformational changes observed in a β6 segment by comparison of free and complexed SBA structures

By comparing the 3-D coordinates of the free enzyme (1BYA structure) with those of the enzyme–substrate complexes (1BYB, 1BYC, and 1BYD), Mikami et al. (1994) observed a strong conformational change in loop 3 (residues 96–103). This change does not take place in the enzyme-inhibitor complex (1BTC), which the authors attributed to steric effects (Mikami et al., 1994). In a recent article (Pujadas et al., 1996), we demonstrated that loop 3 in β-amylase displays two motifs, both of which are highly conserved throughout the evolution and which form a part of a fingerprint. This fingerprint is defined as a set of conserved sequence patterns (Attwood & Beck, 1994).

In the present paper, we describe that by carefully superposing the coordinates from the five reported SBA structures we were able to detect a second conformational transition, which affects both the enzyme-substrate and the enzyme-inhibitor complexes in the same way when the ligand binds to the enzyme. The transition affects three β6 residues ($^{341}FTC^{343}$) and entails an overall change in the geometry of the C-terminal part of the barrel (caused by a displacement of their main chain towards the interior of the barrel; see Kinemage 1). The side chains of these three residues also showed conformational changes, which are considerable in the case of Thr_{342} and Cys_{343} (Kinemage 2). The $^{341}FTC^{343}$ segment is part of the highly conserved motif located in the β6/L6 interface that constitutes the β-amylase fingerprint (Pujadas et. al., 1996).

For the sake of clarity, Figure 1 shows the six CA-layers that make up the SBA hyperboloid scaffold, using the definitions reported elsewhere (Pujadas et al., 1996). Layers 3, 4, and 5 (displayed on a gray background in the figure) are made up of residues, the side chains of which are packed in the central region of the barrel (Pujadas et al., 1996). In addition, layers 5 and 6 (highlighted within a second rectangle in the figure) are made up of eight invariable residues, if alignments are made for all the β-amylase sequences reported up to now (Pujadas et al., 1996). Thr_{342}, a residue involved in substrate contacts (Mikami et al., 1994), belongs to layer 6, whereas Phe_{341} and Cys_{343} are located in the outer part of the barrel flanking Thr_{342}.

Analysis of the crystallographic quality of the segment affected by the conformational changes ($^{341}FTC^{343}$)

To ensure the correctness of the crystallographic assignment in the segment affected by the conformational transition, the five complete SBA structures were analyzed with the Biotech routines (Vriend, 1990; Laskowski et al., 1993; Wodak et al., 1995). No significant mistakes were found when the quality of the $^{341}FTC^{343}$ structures were analyzed. All $^{341}FTC^{343}$ fragments have normal side-chain conformational parameters (unusual rotamers were not found when WHAT_IF was applied; Vriend, 1990), although in some structures the χ_1 dihedral angle of Thr_{342} (1BYA, 1BYC and 1BYD) and Cys_{343} (1BYA) deviates by more than one standard deviation from the mean value (three in the case of Thr_{342} in 1BYA, two for Cys_{343}, which is also in 1BYA, and one for the rest). At the main-chain level, WHAT_IF detected an unusual pair of ϕ/ψ values for Thr_{342} in 1BYA (a detailed explanation of this finding will be given below). Within the $^{341}FTC^{343}$ fragment, Biotech routines also found that some of the interatomic distances were shorter than normal. Of these distances, the most significant are: Phe_{341} and Asn_{340} (N-ND2 in 1BYD), Phe_{341} and Thr_{342} (CG-N in 1BYA), Phe_{341} and Gly_{379} (O-CA in 1BYA and 1BYD), Thr_{342} and Cys_{343} (C-SG, O-SG and CB-N in 1BYA), Thr_{342} and H_2O 505 (CB-HOH, CA-HOH, C-HOH, O-HOH in 1BYD), Thr_{342} and H_2O 784 (CG2-HOH in 1BYA), Cys_{343} and Gly_{298} (SG-CA in 1BYA), Cys_{343} and H_2O 507 (SG-HOH in 1BYB), Cys_{343} and H_2O 744 (SG-HOH in 1BYB), Cys_{343} and SEO_{343} (SG-OB1 and SG-SG in 1BTC). All five structures showed other minor deviations from ideal values in main-chain bond lengths, bond angles, and atom volumes. From our quality analysis we conclude that the transition

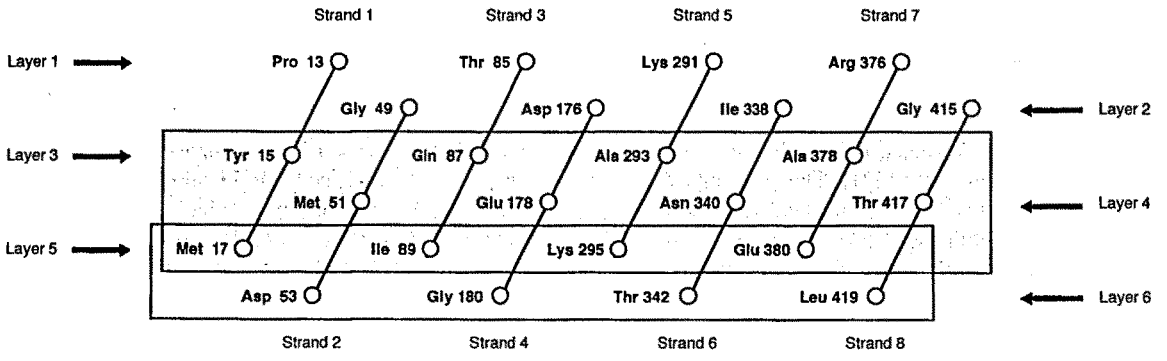


Fig. 1. CA-layered residues located inside the β-barrel in the crystallized SBA. Layer 4 is on the central plane of the hyperboloid. The layers inside the gray rectangle are responsible for the side-chain packing in the interior of the barrel (layers 3, 4, and 5). Layers 5 and 6 inside the second rectangle have residues with non-carbon side-chain atoms located at distances less than 3 Å from the substrate in 1BYB (Asp 53, Lys 295, Thr 342, and Glu 380). Residues from the β-amylases, which are known at present and which are assigned by multialignment to form the putative layers 5 and 6, are conserved in all sequences.

of the ^{341}FTC^{343} fragment should be considered to be a phenomenon that is associated with ligand-binding effects and, therefore, free of any experimental suspicion.

Detailed analysis of the conformational transition associated to the ligand binding in ^{341}FTC^{343} structure fragment

As shown in Table 1, the distance between the opposite Asp_{53} and Thr_{342} CA atoms decreases from 16.5 Å (the enzyme value) to 14.8 ± 0.0 Å (the average value for the complexes). It should be pointed out that all the opposite CA distances for layers 1 to 5 are highly preserved when the ligand binds on. Similarly, the angle of reference chosen for the pair of distance segments increases significatively for layer 6: from 69.4° to 73.4 ± 0.6°.

The Ramachandran angles of the four residues in β6 that precede residues 341–343 undergo no transition (Table 2) when the ligand binds on. In contrast, Figure 2 shows clear-cut transitions of the dihedral angles ϕ and ψ for residues 341–343. As can be observed in Figure 2, Thr_{342} undergoes a considerable transition, changing from a forced β position in the Ramachandran space (ϕ: −143.4°; ψ: −160.7°, for the free enzyme) to a position near the typical right-handed α-helix subspace (ϕ: −77.9 ± 4.6°; ψ: 19.2 ± 6.9°, in complexes). The main-chain conformation of Cys_{343} also changes quite considerably. The Cys_{343} angles in the left-handed α-helix subspace (ϕ: 52.8°; ψ: 36.0°, for the free enzyme) change to more characteristic β-chain values (ϕ: −148.5 ± 4.3°; ψ: 114.6 ± 5.5°, for the complexes). Although notable changes in backbone angles are also observed for Phe_{341}, the transition remains within the β-strand subspace (ϕ moves from −103.1° to −89.5 ± 3.2°; and ψ moves from 179.3° to 111.4 ± 4.0°).

In general, dihedral angles for rotamers throughout β6 appear to be somewhat more permissive than dihedral angles for backbone (Table 2). In the case of Phe_{341}, there is an observable change in χ_2 (going from 45.7° to 61.9 ± 3.2°). In our β-strand specialized rotamer library (Table 3), trans/90 for Phe represents 22.0% of all Phe rotamers, the mean value for χ_1 being −178.1 ± 9.5° and χ_2 being 75.1 ± 20.8°. Two points should be mentioned: (a) the Phe rotamer χ_2 angle for the free enzyme is a long way from the mean value but within the χ_2 experimental distribution limits; the lower and upper limit values being 12.6 and 134.4°, respectively (results not shown); and, (b) in the presence of a ligand, the χ_2 angle increases to values of 61.9 ± 3.2°, thus approaching the mean value in the library. The next two residues in the β chain, Thr_{342} and Cys_{343}, have χ_1 values for the isolated enzyme (−14.3° and 28.9°, respectively) that are highly atypical in β-strand structures. The value for Thr_{342} falls on the limit of its g^+ experimental distribution (mean value and standard deviation: −60.2 ± 9.6° (Table 3); range: from −105.9 to −13.8° (results not shown); frequency in β-strand structures: 54.4%), but Cys_{343} falls outside all the distributions (the nearest is the g^- conformation with lower and upper experimental limits of 51.3 and 77.0° (results not shown)) and is included inside the "Other" category (Table 3) with a frequency in β-strands of 0.6%. After the ligand binds to the enzyme, the values dramatically change to 78.3 ± 6.9° (g^-) and 178.4 ± 5.9° (t)

Table 1. *Distances (in Å) between pairs of layers forming CA-atoms from opposite β-strands (first and second row for each layer) and angles formed between the corresponding distance segments (third row for each layer)*[a]

		1bya	1btc	1byb	1byc	1byd	Average [b]
Layer 1	Pro 13–Lys 291	14.6	14.5	14.5	14.6	14.6	14.6 ± 0.1
	Thr 85–Arg 376	18.7	18.5	18.5	18.5	18.5	18.5 ± 0.0
	Angle	76.2	75.9	75.6	76.3	76.8	76.1 ± 0.5
Layer 2	Gly 49–Ile 338	16.9	16.9	16.7	16.8	16.7	16.8 ± 0.1
	Asp 176–Gly 415	11.0	10.8	10.9	10.8	10.8	10.8 ± 0.1
	Angle	94.0	92.6	93.1	93.1	94.1	93.2 ± 0.7
Layer 3	Tyr 15–Ala 293	14.4	14.4	14.4	14.3	14.3	14.3 ± 0.1
	Gln 87–Ala 378	12.2	11.9	12.0	12.1	12.0	12.0 ± 0.1
	Angle	80.0	79.7	79.2	79.0	79.7	79.4 ± 0.4
Layer 4	Met 51–Asn 340	12.5	12.4	12.5	12.5	12.6	12.5 ± 0.1
	Glu 178–Thr 417	11.9	11.9	11.9	11.9	11.8	11.9 ± 0.1
	Angle	98.1	98.5	99.1	99.0	99.2	98.9 ± 0.3
Layer 5	Met 17–Lys 295	14.2	14.1	14.2	14.2	14.2	14.2 ± 0.0
	Ile 89–Glu 380	13.8	13.6	13.6	13.8	13.7	13.7 ± 0.1
	Angle	92.7	92.0	92.2	92.0	92.0	92.1 ± 0.1
Layer 6[c]	Asp 53–Thr 342	16.5	14.8	14.8	14.9	14.9	14.8 ± 0.0
	Gly 180–Leu 419	16.9	16.8	17.0	16.9	17.1	16.9 ± 0.1
	Angle	69.4	72.7	73.2	73.6	74.1	73.4 ± 0.6

[a] All values are shown for the five β-amylase structures available at present in PDB.
[b] The average values have been calculated from the complexed enzyme structures to facilitate comparison with the free enzyme structure.
[c] As seen from distance and angle data, the conformational change of part of β6 affects not only the CA–CA distances in layer 6 but also the relative position of their CA–CA pairs.

Table 2. *Main- and side-chain dihedral angles (in degrees) of residues that form β-strand six*[a]

		1bya	1byb	1byc	1byd	1btc	Average[b]
				Values for the main-chain angles			
Ala 337	ϕ	−96.1	−97.4	−97.6	−98.2	−102.3	−98.9 ± 2.3
	ψ	160.6	160.0	163.1	161.7	162.1	161.7 ± 1.3
Ile 338	ϕ	−97.9	−97.3	−95.5	−98.9	−94.2	−96.5 ± 2.1
	ψ	139.2	141.0	139.5	140.3	140.6	140.4 ± 0.6
Leu 339	ϕ	−95.6	−97.1	−95.3	−99.0	−96.7	−97.0 ± 1.5
	ψ	110.6	107.0	108.6	110.6	108.7	108.7 ± 1.5
Asn 340	ϕ	−100.8	−95.3	−95.3	−94.9	−92.8	−94.6 ± 1.2
	ψ	107.9	104.2	104.8	105.9	107.6	105.6 ± 1.5
				Values for the side-chain dihedral angles			
Ile 338	χ_1	−70.2	−63.4	−64.5	−66.6	−60.4	−63.7 ± 2.6
	χ_2	174.8	166.4	171.8	175.1	169.3	170.7 ± 3.7
Leu 339	χ_1	−154.5	−166.6	−163.3	−172.7	−163.5	−166.5 ± 4.4
	χ_2	−176.7	168.6	166.7	165.9	156.2	164.3 ± 5.6
Asn 340	χ_1	−178.0	173.5	178.7	175.7	−178.6	177.3 ± 3.5
	χ_2	−169.7	−176.7	−179.1	−179.4	−170.4	−176.4 ± 4.2
Phe 341	χ_1	−161.1	−170.2	−167.8	−166.8	−169.1	−168.5 ± 1.5
	χ_2	45.7	66.6	61.5	59.6	59.9	61.9 ± 3.2
Thr 342	χ_1	−14.3	78.5	82.3	83.9	68.6	78.3 ± 6.9
Cys 343	χ_1	28.9	176.6	174.5	175.4	−172.8	178.4 ± 5.9

[a]Values are indicated for the five β-amylase structures available at present in PDB.
[b]The average values have been calculated from the complexed enzyme structures to facilitate comparison with the free enzyme structure. Main-chain dihedral angles for residues from 341 to 343 are shown in Fig. 2.

Table 3. *Extract from the β-strand rotamer library*[a]

	Rotamer label		Residues in rotamer		Mean value (standard deviation)	
Residue	χ_1	χ_2	Number	%	χ_1	χ_2
Cys	t		52	33.1	−176.9 (10.7)	
	g^+		78	49.7	−63.0 (10.6)	
	g^-		26	16.6	64.7 (7.2)	
	Other		1	0.6		
His	t	t	13	5.3	−177.9 (7.7)	−179.6 (13.6)
	t	−90	38	15.4	−178.0 (13.7)	−102.1 (21.9)
	t	90	39	15.8	−179.9 (10.7)	76.9 (20.1)
	g^+	t	14	5.7	−66.8 (5.6)	169.8 (13.5)
	g^+	−90	73	29.6	−64.1 (11.5)	−84.8 (23.4)
	g^+	90	26	10.5	−65.3 (11.5)	90.6 (22.1)
	g^-	t	2	0.8	62.1 (1.1)	−173.9 (12.3)
	g^-	−90	23	9.3	59.7 (11.0)	−92.9 (17.7)
	g^-	90	17	6.9	61.8 (9.4)	87.7 (12.4)
	Other		2	0.8		
Phe	t	90	119	22.0	−178.1 (9.5)	75.1 (20.8)
	g^+	90	301	55.6	−65.6 (9.6)	94.4 (24.6)
	g^-	90	120	22.2	62.9 (9.2)	92.7 (9.1)
	Other		1	0.2		
Thr	t		90	12.3	−171.4 (9.8)	
	g^+		398	54.4	−60.2 (9.6)	
	g^-		237	32.4	60.7 (10.1)	
	Other		7	1.0		

[a]This table is an extract limited to Cys, Phe, Thr (the three residue types involved in main and side chain conformational transitions in β-amylase), and His (involved in rotamer transitions in endoglucanase CelC). The full β-strand specific rotamer library showing rotamers for all amino acids is in the table in the Electronic Appendix.

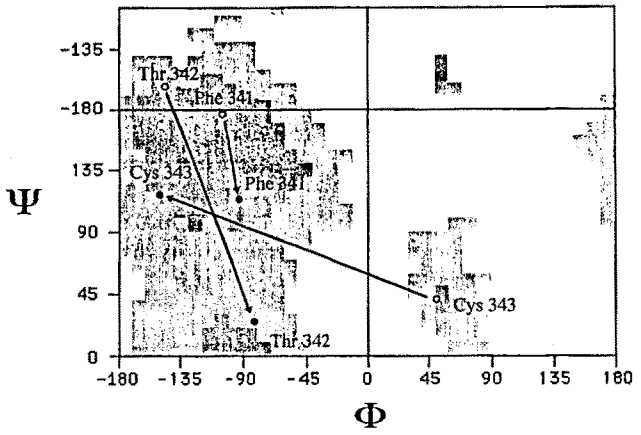


Fig. 2. Ramachandran plot of the 341–343 segment of the SBA sequence. Positions at the start of the arrow (in white) indicate main-chain dihedral angles for the free enzyme structure (1BYA) (Phe 341: $\phi = -103.1°$, $\psi = 179.3°$; Thr 342: $\phi = -143.4°$, $\psi = -160.7°$; Cys 343: $\phi = 52.8°$, $\psi = 36.0°$), whereas the end points (in black) indicate the mean dihedral angles values for the complexed structures (1BYB, 1BYC, 1BYD, and 1BTC) (Phe 341: $\phi = -89.5° \pm 3.2$, $\psi = 111.4° \pm 4.0$; Thr 342: $\phi = -77.9° \pm 4.6$, $\psi = 19.2° \pm 6.9$; Cys 343: $\phi = -148.5° \pm 4.3$, $\psi = 114.6° \pm 5.5$).

for Thr_{342} and Cys_{343} respectively. Their corresponding mean values in the library (Table 3) are 60.7 ± 10.1° (32.4% in β-strands) and −176.9 ± 10.7° (33.1% in β-strands). The χ_1 experimental distribution ranges for g^- and t in Thr and Cys are from 12.2 to 90.9° and from 159.4 to −142.6°, respectively (results not shown).

Analysis of the interactions involving residues $^{341}FTC^{343}$ in the free and ligand-bound enzymes

For the free enzyme structure (1BYA), several interactions appear to be important: The -SH group of Cys_{343} and the main-chain carbonyl oxygen of Thr_{342} are located at a distance of 2.38 Å, with a strong hydrogen bond between them (Donor-Hydrogen-Acceptor (DHA) angle: 175.3°); the aromatic ring of Phe_{341} adapts itself onto Cys_{343} SG and Thr_{342} O; Thr_{342} OG1 and the water molecule HOH_{828} form a strong hydrogen bond (distance: 2.48 Å; DHA angle: 124.3°); two additional hydrogen bonds, both of which have Phe_{341} and two β7-strand residues, are detected between Phe_{341} N and Ala_{378} O (distance: 2.71 Å; DHA angle: 166.9°) and between Glu_{380} N and Phe_{341} O (distance: 2.58 Å; DHA angle: 130.0°).

For the enzyme–ligand structures (1BYB, 1BYC, 1BYD, and 1BTC), several remarks should be made: (a) the hydrogen bond between Cys_{343} SG and Thr_{342} O no longer exists; (b) in one case (1BYB), the Cys_{343} SG interacts with two water molecules, HOH_{507} and HOH_{744}, at 2.85 and 2.69 Å respectively (however, for 1BYC, 1BYD and 1BTC no Cys_{343} SG-interacting water molecule was detected); (c) for 1BTC Cys_{343} SG is covalently bridged with SEO_{343} SG (2-mercaptoethanol) at 2.02 Å; (d) the two above-mentioned hydrogen bonds with Phe_{341} and two β7-strand residues are preserved in all cases, with some enlargement in donor–acceptor distances; (e) Thr_{342} O, N, and OG1 are involved, either as donors or as acceptors, in hydrogen bonds formed with substrate atoms, or with water molecules, or with atoms belonging to the enzyme. All interactions described in this section are shown in Kinemage 3.

Energetics of the conformational transition of the $^{341}FTC^{343}$ segment

When the transition takes place from the free enzyme (1BYA) to the substrate-bound enzyme (1BYC), potential energy differences are negative and very similar for the three-residue segment and the whole β6 segment (−27 kcal/mol for $^{341}FTC^{343}$ and −28 kcal/mol for $^{337}AILNFTC^{343}$). When the transition takes place from 1BYA (free enzyme) to 1BTC (enzyme-inhibitor complex) the negative potential energy differences are somewhat higher (−34 kcal/mol for $^{341}FTC^{343}$ and −43 kcal/mol for $^{337}AILNFTC^{343}$). These higher values probably are due to small differences observed in the main-chain and rotamer dihedral angles. The model system chosen for the calculations is purposely simplified and the values for enthalpic variations should be taken as approximate (Kuntz et al., 1994). In any case, the structural transition observed for β6 appears as an exothermic process, in concordance with the more relaxed conformation of the $^{341}FTC^{343}$ segment when the ligand is attached to the enzyme.

Searching for other examples of free and ligand-bound glycosyl hydrolases of a $(\beta/\alpha)_8$-barrel folding type that might show changes in the main- and side-chain dihedral angles of the barrel core

We searched within the SCOP database for other glycosyl hydrolases with $(\beta/\alpha)_8$-barrel folding that could show changes caused by a possible catalytic or ligand-binding mechanism in the main- and side-chain dihedral angles of their β-sheet barrel core. We needed to find structures for a given enzyme (from the same source) that had been crystallized with and without substrates (or inhibitors). Two other enzymes fulfilled such conditions: the endoglucanase CelC (E.C. 3.2.1.4), which belongs to the beta-glycanase family, from *Clostridium thermocellum*, and the hevamine A (E.C. 3.2.1.14/3.2.1.17), which belongs to the type II chitinase family, from *Hevea brasiliensis*. The PDB entries for their structure are 1CEC, 1CEN, and 1CEO (Dominguez et al., 1995), and 1HVM, 1HVQ, and 1LLO (Terwisscha-van-Scheltinga et al., 1995), respectively. In the case of the endoglucanase CelC, the careful superposition of each one of the barrel core β-strands from the three structures revealed no changes in the backbone conformation. With respect to the side chains, a clear rotamer transition occurs for His_{91} and His_{90} (β-strand 3) (Kinemage 4). The binding of cellohexaose to the enzyme makes His_{90} change from $t/90$ to $t/-90$ rotamer (χ_1/χ_2 values for 1CEC, 1CEO, and 1CEN are −173.1/67.0, −171.3/84.4, and −153.7/−87.5°, respectively) and His_{91} change from $g^+/-90$ to $t/90$ rotamer (χ_1/χ_2 values for 1CEC, 1CEO, and 1CEN are −73.2/−55.8, −79.6/−18.1, and 178.6/11.3°, respectively). It has recently been reported (Dominguez et al., 1996) that His_{90} is one of the seven amino acid residues (the others are Arg_{46}, Asn_{139}, Glu_{140}, His_{198}, Tyr_{200}, and Glu_{280}) that are located close to the catalytic reaction center and strictly conserved in family 5 cellulases. Moreover, His_{90} participates in the catalytical process by making direct contact with the substrate and forming part of a network of hydrogen bonds. This network contributes to the stability of the active site architecture and may influence the protonation state of the two catalytic residues (Dominguez et al., 1996).

The same analytical approach was applied to the hevamine A structures but neither backbone nor side-chain conformational changes were detected in the barrel forming β-strands.

Discussion

The $^{341}FTC^{343}$ segment is shown to act as a "trap trigger" in the binding process of the ligand to the enzyme

To our knowledge, this is the first report that describes considerable conformational changes in both the main and the side chains in the barrel β-sheet core of a $(\beta/\alpha)_8$-barrel glycosyl hydrolase after substrate (or inhibitor) binding to the enzyme. The conformational transition involves only three residues (from 341 to 343 in the crystallized SBA), Thr_{342} being a residue that is part of the C-terminal CA-layer of the barrel (Fig. 1; Pujadas et al., 1996). These three residues are conserved in most known β-amylase sequences (21 β-amylase protein sequences), the only exception being the sequence for *Bacillus circulans* β-amylase (BCAMYB), which has an Ala instead of a Cys in the position equivalent to SBA Cys_{343}. The conformational effect due to the binding on of the ligand is almost identical whether the molecule is a substrate or an inhibitor. This indicates that the transition is related to the attachment mechanism of the enzyme. The region of the molecule where the phenomenon takes place (Fig. 1) is made up of two CA-layers (layers 5 and 6) consisting of eight residues that undergo no change throughout the evolution, either in plants or in microorganisms (Pujadas et al., 1996). These two layers are involved in the substrate contacts (e.g., Thr_{342}) and also, apparently, in the enzymatic activity (Mikami et al., 1994). Within β6, the neighbors Phe_{341} and Cys_{343} occupy an external position in the barrel.

The transition caused by the binding of the ligand (clearly observed in Fig. 3) drastically alters the backbone and side-chain dihedral angles in such a way that several atoms draw nearer to the ligand molecule; e.g., OG1 of Thr_{342} is located at 2.8–2.9 Å from the O3 and 3.1–3.3 Å from the O4 atoms of the glucose molecule that is labeled as GLC 498 in 1BYB and 1BYC or DOM 498 in 1BYD (Kinemage 3). In this case, Thr_{342} follows a tendency also observed in small polar residues located in disallowed or unusual Ramachandran conformations where backbone distortions are compensated by hydrogen bond interactions involving functional side chains (Gunasekaran et al., 1996). Table 1 shows also a significative shortening of the CA distance in layer 6 between Asp_{53} and Thr_{342} (from 16.5 Å to 14.8 ± 0.0 Å).

The net changes in ϕ and ψ angles (Figs. 2, 3) for the three residues and their unrelaxed χ values for the free enzyme (in comparison with the more "normal" values when the ligand is attached) provide strong evidence that the small segment formed by these three residues acts as a "trap trigger" to adapt the ligand to the active site. In the case of the substrate, loop 3 simultaneously approaches the active center, acting as a hinged lid (Mikami et al., 1994). Therefore, the $^{341}FTC^{343}$ segment in SBA should be an interesting target for studying enzyme activity based on site-directed mutagenesis.

Local steric strain of the free enzyme, present in the $^{341}FTC^{343}$ backbone and their side chains, may facilitate the binding to ligands and catalysis

In some proteins, a local strain in the polypeptide backbone has been identified as being necessary for ligand binding and catalysis (Herzberg & Moult, 1991). This strain is reflected as a set of ϕ, ψ values that act as outliers in conflictive regions of the Ramachandran map (despite the fact that the analyzed structures showed very good quality parameters). It has been argued by Herzberg and Moult (1991) that this local strain may be needed for the precise distribution of the functional groups allowing the binding and/or the catalytical mechanism.

We explain our observations of "rare transitions" for Thr_{342} and Cys_{343} in the Ramachandran map using Herzberg and Moult's (1991) proposal that there are few sterically strained features and these are overwhelmingly located in regions concerned with function. In the case of Thr_{342}, the transition from a forced canonical to a non-canonical β-strand Ramachandran position after ligand binding (Fig. 2) brings this residue nearer to the substrate (Kinemage 3). In the case of Cys_{343}, the transition from a backbone strain in the free enzyme to a more relaxed β-strand position may be necessary for the "trap trigger" action (Fig. 2). The Phe_{341} is the only residue of the $^{341}FTC^{343}$ segment with no changes in the backbone steric strain or in the rotamer. Consequently, the changes in this residue may be due to an adaptation to the new conformational constraints imposed by $^{342}TC^{343}$ after ligand binding. In any case, such an adaptation also contributes to the enthalpic variations.

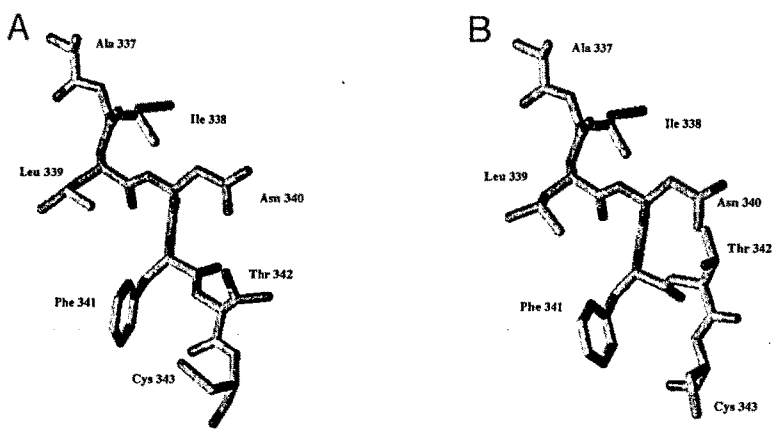


Fig. 3. A diagram of conformational change. **(A)** belongs to the 1BYA (free enzyme) and **(B)** belongs to 1BYB (enzyme complexed with maltose). The conformational changes not only affect the main chain of β-strand six (Fig. 2), but also some of the side chains (Table 2). The conformation of this strand in the complexed structures (1BYB, 1BYC, 1BYD, and 1BTC) is equivalent. The pictures have been generated using the program MolView v1.4.1 (Smith, 1995).

Shoichet et al. (1995) have proposed a "stability-function" hypothesis, which states that a balance between the stability and the function of enzymes can be obtained so that they use their ordered structures to facilitate catalysis. In addition to their own results on the stability and activity of T4 lysozymes with substitutions in the substrate-binding site, some of Shoichet's findings derive from earlier studies by Warshel et al. (1988) and Meiering et al. (1992). In the case of β-amylase, the enthalpic cost of the local conformational transition that involves the ^{341}FTC^{343} segment is negative, which is strong evidence that the active-site cleft that includes this segment is of a metastable nature. In contrast to the tendency of global folding, which minimizes the free energy of any enzyme, a local surplus of macromolecule instability may maximize the functional role of the active site. Other mechanisms involving conformational changes in the active site of proteins will probably be described in the future. In this respect, a recent paper by Skinner et al. (1997) provides insight into the change in binding affinity of the heparin pentasaccharide that accompanies a conformational change of intact antithrombin, the highest resolution structure of an active inhibitory serpin to date.

A postulated catalytical role for the ^{341}FTC^{343} segment

The molecular cloning and sequencing of *B. circulans* β-amylase (Siggens, 1987) revealed the presence of Ala_{359} instead of Cys_{359} (which is the equivalent of Cys_{343} in soybean). This is the only case reported in the bibliography or in sequence databases in which such a crucial change occurs in a living cell. The expression of the isolated gene in *Escherichia coli* NM522 revealed a high production of the enzyme, which is probably located in the periplasmic space but not secreted into the culture supernatant. The β-amylase activity in cell lysates was analyzed using SDS-PAGE and then a blot was made onto a gel containing starch. However, enzymatic activity should be quantified and compared to other Cys- (not Ala-) containing β-amylases from related microorganisms (e.g., *Bacillus polymyxa*) if the meaning of the spontaneous mutation in terms of catalysis is to be clarified.

The mutation Cys^{343}Ser, obtained by directed mutagenesis, reduces the native activity in SBA by two-thirds (Totsuka et al., 1994) and the equivalent mutation (Cys^{323}Ser) reduces the activity in *B. polymyxa* β-amylase by five (Uozumi et al., 1991). These later authors considered that the thiol hydrogen of Cys_{323} might participate in the formation of a hydrogen bond that, if absent, subtly alters the enzyme structure and lowers the specific activity of the Cys^{323}Ser *B. polymyxa* mutant. Nomura et al. (1995) understood from these findings—applied to *Bacillus cereus*—that Cys_{331} is indispensable for stabilizing Thr_{330}, the OG1 atom of which forms hydrogen bonds with hydroxyl groups O-3 and O-4 of glucose residues at subsite 3. In this interpretation, the existence in SBA of a local conformational change involving Cys_{343}, in particular, and the segment ^{341}FTC^{343} as a whole, would favor the spatial interplay of Thr and its environment within the active cleft. According to a proposed action mechanism, the water molecule HOH_{866}, which is hydrogen bonded between Glu_{380} and Thr_{342}, may attack the anomeric carbon, reverse the configuration, and produce β-maltose (Totsuka & Fukazawa, 1996); in this respect, local conformational changes in the short segment where Thr_{342} is inserted may facilitate the way for a back-and-forth repetitive mechanism.

The preferred side-chain conformations of cysteinyl residues in helices are the ones that potentially form intrahelical hydrogen bonds with carbonyl oxygens in the preceding turn of the helix (Gray & Matthews, 1984). Using the HBPLUS program (McDonald & Thornton, 1994), 46 hydrogen bonds (with Cys SG acting as a donor and the oxygen of the main chain as an acceptor) were found in more than 18,000 β-strands from 1,700 highly resolved proteins present in the PDB; the great majority of distances ranged from 3.4 to 3.8 Å (results not shown). These findings are in agreement with the equilibrium distance of 3.66 Å from donor to acceptor in a hydrogen bond between a sulfhydryl group and oxygen (Gregoret et al., 1991). However, our search for interactions involving atoms of the segment ^{341}FTC^{343} reveals that there is a short (2.38 Å), very strong, low-barrier hydrogen bond between Cys_{343} SG and Thr_{342} O (the main-chain carbonyl oxygen) in the free enzyme. This is the first report of a low-barrier hydrogen bond involving a sulfhydryl group. It is interesting to note that the formation of such a bond can supply 10–20 kcal/mol (Cleland & Kreevoy, 1994).

A second stabilizing element of the ^{341}FTC^{343} segment in the free enzyme is the Phe_{341} side chain, which is adjacent to the Cys_{343} SG and Thr_{342} O nucleophiles. These two nucleophiles perturb the π electron system of the aromatic ring. The nucleophilic effect of these two atoms is facilitated by the DHA angle that they form and which is almost planar (175.3°). Viguera & Serrano (1995) have pointed out that there is a strong electrostatic interaction between the electronegative sulfur atom and the positively charged aromatic ring hydrogens. In our case, the existence near the Phe_{341} ring of a second nucleophile, such as Thr_{342} O, should enhance the electrostatic effect. The formation of such an interacting structure may supply additional energy to the ^{341}FTC^{343} segment.

The loss of β-amylase activity in *B. polymyxa* for the mutated Cys^{323}Ser enzyme variant (Uozumi et al., 1991)—and, to a lesser extent, of SBA activity (Totsuka et al., 1994)—may be because the hydrogen bond that is formed between Ser OG and Thr O (Ser_{323} OG and Thr_{322} O for *B. polymyxa*) has a different potential function. In order to understand the functioning of the active cleft of β-amylase better, both single and double mutants should be formed by combining mutations such as Cys^{343}Ser, Cys^{343}Thr, Thr^{342}Ser, and Phe^{341}Tyr for SBA or in equivalent positions for the corresponding sequences for other species.

It can be concluded that the low-barrier hydrogen bond between Cys_{343} SG and Thr_{342} O is essential if the enzyme is to remain in its native and metastable form while waiting for the ^{341}FTC^{343} segment to adopt a more relaxed form when the substrate is bound to the enzyme. Generally speaking, this is the way our postulated "trap trigger" would function within the active cleft.

Final remarks

At present, the SBA is the only known crystallized E.C. 3.2.1.x enzyme (with $(\beta/\alpha)_8$-barrel folding) with considerable changes in both the main- and side-chain dihedral angles of the β-sheet barrel core during ligand binding. Only two other glycosyl hydrolases (with the same folding) are known to be crystallized, from a single source, in the absence and in the presence of a ligand and they show either limited rotamer transitions (endoglucanase CelC) or no conformational variation (hevamine A) in their barrel core.

In this work, we provide computational evidence for the SBA, which shows the metastable conformation of a short sequence segment in the active-site cleft for the free enzyme and a more relaxed conformational arrangement of the same segment for the ligand-bound enzyme. In accordance with our results, we propose

an all-or-none-effect mechanism (the "trap trigger" mechanism) to adapt the ligand to the catalytical site. Glycosyl hydrolases with a TIM-barrel-like catalytical domain might commonly use conformational transitions similar to the ones we found in the β6 strand of β-amylase to degrade polysaccharides. Whether such results can be extrapolated to other glycosyl hydrolases that share this folding is an issue that, in our opinion, merits further experimental attention. However, before such computational work can be done, more examples of both free and ligand-bound glycosyl hydrolases should be resolved by crystallographers.

Our proposal that β-amylase has a "trap trigger" mechanism that makes it easier for the ligand to adapt to the active sites supports the general hypothesis for enzymes postulated by Herzberg and Moult (1991) and by Shoichet et al. (1995).

Materials and methods

Three-dimensional data for SBA (Mikami et al., 1993; Mikami et al., 1994) were imported from the last on-line release of PDB (Bernstein et al., 1977) (http://www.pdb.bnl.gov/cgi-bin/browse) using the Netscape navigator application v3.0 Gold installed in a Macintosh personal computer and connected to the WWW server via the Internet. The PDB entries found were 1BYA, 1BYB, 1BYC, 1BYD, and 1BTC; resolution: 2.2, 1.9, 2.2, 2.2, and 2.0 Å, respectively; *R*-factor: 0.169, 0.149, 0.149, 0.145, and 0.177, respectively. The enzyme was found to be free (1BYA) or to form a complex with substrates (1BYB, 1BYC, 1BYD) or an inhibitor (1BTC). The sequences for the five structures were identical. The current on-line version of SCOP (http://scop.mrc-lmb.cam.ac.uk/scop/) (Murzin et al., 1995) was used to search for other glycosyl hydrolases with a TIM-barrel-like structure that are crystallized in the absence and in the presence of ligands from a single source.

Sets of coordinates that best fitted the five 3-D structures were obtained by using the ASSEMBLING program, written in FORTRAN 77 (G. Pujadas & J. Palau, unpubl.) and were merged in one file so that the conformational changes and water movements between structures could be easily identified on the same computer screen. Backbone and side-chain traces, for the whole chain, were inspected for variations with the RasMol program (v2.6, Glaxo Research and Development, Greenford, Middlesex, United Kingdom) (Sayle & Milner-White, 1995), using a Silicon Graphics Indigo^2 XZ. The conformational changes were identified by comparing superposed structure segments of about 50 residues from the whole set of homologous sequences. Relative distances that might reveal interactions in which atoms of the ^{341}FTC^{343} segment participate were also examined with RasMol. Interactions qualified as hydrogen bonds with acceptable geometries were determined with the HBPLUS program, v3.15 (McDonald & Thornton, 1994).

The distances between the two pairs of opposite CA atoms in a layer and the angle between the two defining segments were measured for the six CA-layers that constitute the SBA $(\beta/\alpha)_8$ scaffold by using the atom coordinates provided by the respective PDB files in an Excel v4.0 worksheet. The dihedral angles for both backbone and side chains were obtained from the *measures* function of MAGE v4.2 (Richardson & Richardson, 1992). The atoms used to calculate the different dihedral angles were the ones recommended by the IUPAC-IUB (IUPAC-IUB, 1970).

In order to ensure that the conformational change in β6 was not an experimental artifact, the quality of the complete set of β-amylase structures, with special attention to the correctness of the ^{341}FTC^{343} segment, was tested with the "Biotech Validation Suite for Protein Structures" (http://biotech.embl-ebi.ac.uk:8400/). All the tools available in the Biotech analysis routines were used ("long check" option). The Biotech criteria compare the geometric characteristics (main-chain covalent bond lengths and bond angles, planarity of aromatic groups and sp^2 side-chain end-groups, main and side-chain dihedral angles, non-bonded interactions, main-chain hydrogen bonds, etc.) of the query structures with standard values derived from high quality and reliable structures. The above analysis routines are part of Prochek v.3.0 (Laskowski et al., 1993), WHAT IF v4.99 (Vriend, 1990), and SurVol (Wodak et al., 1995), and are implemented in the Biotech WWW server.

Rotamers were assigned by comparing the side-chain dihedral angles for Phe_{341}, Thr_{342}, and Cys_{343} in SBA, and His_{90} and His_{91} in endoglucanase CelC with our current β-strand rotamer library (Table 3 for Cys, His, Phe and Thr and the Electronic Appendix Table for the complete set of amino acids). This library was derived from a high quality, low sequence redundance set of 220 protein subunits (resolution and *R*-value less than or equal to 2.0 Å and 0.200; the average maximum homology allowed was 35%) to guarantee the correctness of the rotamer distribution and prevent sample overrepresentation. This set was obtained from the last 1996 PDB_Select release (pdb_select.1996-Nov-11) (Hobohm et al., 1992; Hobohm & Sander, 1994), downloaded via anonymous ftp from the EMBL server (ftp.embl-heidelberg.de/pub/databases/pdb_select), and contained 10,335 residues (157 Cys, 247 His, 541 Phe, and 732 Thr) in 1803 β-strands. In the case of Phe (and also Tyr (the Electronic Appendix Table)), the atom labels of the aromatic ring were corrected, if necessary, to comply with rule 2.3.2 of the IUPAC-IUB Commission (IUPAC-IUB, 1970). The nomenclature for χ_1 is consistent with the one used by Janin et al. (1978), McGregor et al. (1987), and Blaber et al. (1994), who define trans (t), gauche$^-$ (g^-), and gauche$^+$ (g^+) as dihedral angle distributions with mean values around 180°, 60°, and −60° (or 300°), respectively (the nomenclature of g^- and g^+ is interchanged with respect to the definition of Ponder & Richards (1987)). The limits for the different conformations of the dihedral angles, which were used to separate coordinates into their rotamers, were obtained from the graphical distribution of their values without taking into account the secondary structure of origin (results not shown). The χ_1 limits used were (a) Cys: t, 140/−130°; g^+, −100/−20°; g^-, 30/100°; (b) His: t, 150/−120°; g^+, −110/−30°; g^-, 20/90°; (c) Phe: t, 150/−130°; g^+, −120/−30°; g^-, 30/100°; and (d) Thr: t, 150/−140°; g^+, −130/−10°; g^-, 10/100°. The χ_2 limits for Phe are 0/180°, whereas for His the values are 150/−150° (t), −150/−10° (−90), and 30/150° (+90), respectively. It should be pointed out that these limits were only used to distribute the coordinates into the different rotamers while the library was being built, but the Results and Discussion sections refer to experimental limit values found from analyzing all the specific rotamer distributions in β-strands.

Potential energy calculations were performed with the GROMOS package (van Gunsteren & Berendsen, 1987) using coordinates taken from the X-ray structures (1BYA, 1BYC, and 1BTC). A few water molecules located in the environment, as well as the substrate or inhibitor molecules, were not considered in our calculations. The potential energy of the short segments of the proteins (^{341}FTC^{343} and ^{337}AILNFTC^{343}) was calculated by descent energy optimization (one step; NIS field; no SHAKE algorithm used). This potential energy accounts for the contributions of the vibrational frequencies from the stretching and bending modes of each valence bond and internal rotations, as well as for the con-

tributions of noncovalent interactions from the secondary and tertiary structure (Gómez et al., 1995). By measuring energy (or enthalpy) changes or using force field calculations, in principle, energy can be assigned to a given process (Mark & van Gunsteren, 1994). The cost of the potential energy used to carry out the local conformational transition was estimated by subtracting the energies calculated for the enzyme–ligand complexes from the energy of the free enzyme. Energy minimization techniques yield only an approximation of the enthalpy and more often give semiquantitative rankings with uncertainties of ±5 kcal/mol (Kuntz et al., 1994). Since force field calculations were made using quite similar peptides in the two series of protein segments chosen, the enthalpy values obtained are expected to be more reliable (Kuntz et al., 1994).

Acknowledgments

We thank Prof. Enric Querol for computational facilities, Dr. Baldomero Oliva for doing the potential energy calculations, Dr. Thomas J. Smith for providing help in the use of the latest versions of the MolView program, the X-ray crystallographer Dr. Ignacio Fita for a thorough discussion concerning some comments made by one referee of the present paper, and John Bates (English Text & Style Correction Service of Universitat Rovira i Virgili) for his help with writing the manuscript. G.P. has received a fellowship (FI/93-129) from the Catalan Governmental Agency CIRIT (Generalitat de Catalunya). Finally, we thank the universitat Rovira I Virgili for use of facilities, although this research had not been awarded grants by any research-supporting institution.

References

Attwood TK, Beck ME. 1994. PRINTS—A protein motif fingerprint database. *Protein Eng 7*:841–848.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:535–542.

Blaber M, Zhang X, Lindstrom JD, Pepiot SD, Baase WA, Matthews BW. 1994. Determination of α-helix propensity within the context of a folded protein. Sites 44 and 131 in bacteriophage T4 lysozyme. *J Mol Biol 235*:600–624.

Cleland WW, Kreevoy MM. 1994. Low-barrier hydrogen bonds and enzymic catalysis. *Science 264*:1887–1890.

Dominguez R, Souchon H, Lascombe M, Alzari PM. 1996. The crystal structure of a family 5 endoglucanase mutant in complexed and uncomplexed forms reveals an induced fit activation mechanism. *J Mol Biol 257*:1042–1051.

Dominguez R, Souchon H, Spinelli S, Dauter Z, Wilson KS, Chauvaux S, Beguin P, Alzari PM. 1995. A common protein fold and similar active site in two distinct families of beta-glycanases. *Nat Struct Biol 2*:569–576.

Gómez J, Hilser VJ, Xie D, Freire E. 1995. The heat capacity of proteins. *Proteins Struct Funct Genet 22*:404–412.

Gray TM, Matthews BW. 1984. Intrahelical hydrogen bonding of serine, threonine and cysteine residues within α-helices and its relevance to membrane-bound proteins. *J Mol Biol 175*:75–81.

Gregoret LM, Rader SD, Fletterick RJ, Cohen FE. 1991. Hydrogen bonds involving sulfur atoms in proteins. *Proteins 9*:99–107.

Gunasekaran K, Ramakrishnan C, Balaram P. 1996. Disallowed Ramachandran conformations of amino acid residues in protein structures. *J Mol Biol 264*:191–198.

Herzberg O, Moult J. 1991. Analysis of the steric strain in the polypeptide backbone of protein molecules. *Proteins Struct Funct Genet 11*:223–229.

Hobohm U, Sander C. 1994. Enlarged representative set of protein structures. *Protein Sci 3*:522–524.

Hobohm U, Scharf M, Schneider R, Sander C. 1992. Selection of representative protein data sets. *Protein Sci 1*:409–417.

IUPAC-IUB Commission on Biochemical nomenclature. 1970. Abbreviations and symbols for the description of the conformation of polypeptide chains. Tentative rules (1969). *Biochemistry 9*:3471–3479.

Janin J, Wodak S, Levitt M, Maigret B. 1978. Conformation of amino acid side-chains in proteins. *J Mol Biol 125*:357–386.

Kuntz ID, Meng EC, Schoichet BK. 1994. Structure-based molecular design. *Acc Chem Res 27*:117–123.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Cryst 26*:283–291.

Mark AE, van Gunsteren WF. 1994. Decomposition of a free energy of a system in terms of specific interactions. Implications for theoretical and experimental studies. *J Mol Biol 240*:167–176.

McDonald IK, Thornton JM. 1994. Satisfying hydrogen-bonding potential in proteins. *J Mol Biol 238*:777–793.

McGregor MJ, Islam SA, Sternberg MJE. 1987. Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *J Mol Biol 198*:295–310.

Meiering EM, Serrano L, Fersht AR. 1992. Effect of active-site residues in barnase on activity and stability. *J Mol Biol 225*:585–589.

Mikami B, Degano M, Hehre EJ, Sacchettini JC. 1994. Crystal structures of soybean beta-amylase reacted with beta-maltose and maltal: Active-site components and their apparent roles in catalysis. *Biochemistry 33*:7779–7787.

Mikami B, Hehre EJ, Sato M, Katsube Y, Hirose M, Morita Y, Sacchettini JC. 1993. The 2.0-Å resolution structure of soybean beta-amylase complexed with alpha-cyclodextrin. *Biochemistry 32*:6836–6845.

Murzin AG, Brenner SE, Hubbard T, Chothia C. 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol 247*:536–540.

Nanmori T. 1988. Bacterial β-amylases (*B. cereus, B. polymyxa*, etc.). In: The Amylase Research Society of Japan, ed. *Handbook of Amylases and Related Enzymes*. Oxford: Pergamon Press. pp 94–99.

Nomura K, Yoneda I, Nanmori T, Shinke R, Morita Y, Mikami B. 1995. The role of SH and S-S groups in *Bacillus cereus* β-amylase. *J Biochem (Tokyo) 118*:1124–1130.

Ponder JW, Richards FM. 1987. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J Mol Biol 193*:775–791.

Pujadas G, Ramírez FM, Valero R, Palau J. 1996. Evolution of β-amylase: Patterns of variation and conservation in subfamily sequences in relation to parsimony mechanisms. *Proteins Struct Funct Genet 25*:456–472.

Richardson DC, Richardson JS. 1992. The kinemage: A tool for scientific communication. *Protein Sci 1*:3–9.

Sayle R, Milner-White EJ. 1995. RasMol: Biomolecular graphics for all. *Trends Biochem Sci 20*:333–379.

Shoichet BK, Baase WA, Kuroki R, Matthews BW. 1995. A relationship between protein stability and protein function. *Proc Natl Acad Sci USA 92*:452–456.

Siggens KW. 1987. Molecular cloning and characterization of the beta-amylase gene from *Bacillus circulans*. *Mol Microbiol 1*:86–91.

Skinner R, Abrahams JP, Whisstock JC, Lesk AM, Carrell RW, Wardell MR. 1997. The 2.6-angstrom structure of antithrombin indicates a conformational change at the heparin binding site. *J Mol Biol 266*:601–609.

Smith TJ. 1995. MolView: A program for analyzing and displaying atomic structures on the Macintosh personal computer. *J Mol Graphics 13*:122–125.

Terwisscha van Scheltinga AC, Armand S, Kalk KH, Isogai A, Henrissat B, Dijkstra BW. 1995. Stereochemistry of chitin hydrolysis by a plant chitinase/lysozyme and X-ray structure of a complex with allosamidin: Evidence for substrate assisted catalysis. *Biochemistry 34*:15619–15623.

Thoma JA, Spradlin JE, Dygert S. 1971. Plant and animal amylases. In: Boyer PD, ed. *The Enzymes, 3rd ed.* New York: Academic Press, pp 115–189.

Totsuka A, Fukazawa C. 1996. Functional analysis of Glu380 and Leu383 of soybean β-amylase. A proposed action mechanism. *Eur J Biochem 240*:655–659.

Totsuka A, Nong VH, Kadokawa H, Kim CS, Itoh Y, Fukazawa C. 1994. Residues essential for catalytic activity of soybean β-amylase. *Eur J Biochem 221*:649–654.

Uozumi N, Matsuda T, Tsukagoshi N, Udaka S. 1991. Structural and functional roles of cysteine residues in *Bacillus polymyxa* β-amylase. *Biochemistry 89*:4594–4599.

van Gunsteren WF, Berendsen HJC. 1987. Groningen Molecular Simulation (GROMOS) Library Manual. Groningen, The Netherlands: Biomos.

Viguera AR, Serrano L. 1995. Side-chain interactions between sulfur-containing amino acids and phenylalanine in α-helices. *Biochemistry 34*:8771–8779.

Vriend G. 1990. WHAT IF: A molecular modeling and drug design program. *J Mol Graph 8*:52–56.

Warshel A, Sussman F, Hwang JK. 1988. Evaluation of catalytic free energies in genetically modified proteins. *J Mol Biol 201*:139–159.

Wodak SJ, Pontius J, Vaguine A, Richelle J. 1995. Validating protein structures: From consistency checking to quality assessment in making the most of your model. In: Hunter WN, Thornton JM and Bailey S, eds. *Proceedings of the CCP4 Study Weekend 6–7 January 1995*. Daresbury, UK: Chadwick Library, pp 41–51.

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8:
RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

UNIVERSITAT ROVIRA I VIRGILI.

DEPARTAMENT DE BIOQUÍMICA I BIOTECNOLOGIA.

Estudis computacionals sobre β-amilasa, una molècula enzimàtica tipus barril $(\beta/\alpha)_8$: relacions evolutives i transicions estructurals al centre catalític.

Memòria presentada per

GERARD PUJADAS ANGUIANO

per optar al grau de Doctor en Ciències Químiques
sota la direcció del Dr. Jaume Palau Albet.

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8:
RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

Conclusions.

• El multialineament del conjunt de seqüències de les β-amilases permet distinguir —amb claretat— dos grups evolutius, el de les plantes i el dels microorganismes. Aquesta conclusió —obtinguda inicialment a partir dels percentatges d'homologia entre les seqüències i la disposició dels forats (*gaps* en anglès) en el multialineament— es veurà ratificada a nivell estructural amb posterioritat.

• El multialineament del conjunt de seqüències de les β-amilases permet distinguir vuit segments —de longitud igual o superior a sis residus— que presenten —segons els alfabets de Karlin— un nivell elevat de conservació tant de les característiques estructurals com de les químiques. Les seqüències consens de molts d'aquests segments —també anomenades motius— són característiques de les β-amilases. El conjunt d'aquests vuit motius constitueixen l'empremta de seqüència —*sequence fingerprints* en anglès— de la β-amilasa. Es suggereix la importància de poder elaborar una base de dades de motius basada en consideracions taxonòmiques.

• L'ànalisi combinat del multialineament de les seqüències d'ADN i del de les proteïnes, revela la presència de mutacions que alteren localment la pauta de lectura —*frameshift mutation* en anglès— de l'ADN. Aquestes mutacions arriben a afectar al Glu del centre actiu de la seqüència SOYAMYLASE que canvia la posició 185 per la 186. Com que aquest Glu es troba en el llaç L_4 —per tant dins d'una estructura flexible— el canvi de posició manté encara l'activitat enzimàtica.

• L'evolució de la seqüència de la β-amilasa es veu restringida per la necessitat de conservar les zones imprescindibles pel manteniment de l'estructura i l'activitat enzimàtica. En aquestes zones el procés de divergència evolutiva ocorre a una velocitat baixa. Els segments corresponents a les zones no crítiques per a l'estructura o la funció —llaços de l'extrem C-terminal del barril i hèlices α— evolucionen de forma parsimoniosa a una velocitat relativament elevada.

• El barril de la β-amilasa està constituït per sis plans perpendiculars a l'eix del barril. Cadascun d'aquests plans passa —aproximadament— per quatre carbonis α de residus encarats cap a l'interior del barril i que pertanyen a tires β alternades. Entre aquestes capes cal distinguir les que són d'interacció —proporcionen els residus que interaccionen a dins del barril— i les que no ho són (que són fruit de la pròpia geometria del barril). Les capes d'interacció de la β-amilasa són la 3, la 4 i la 5 (on l'ordre creixent indica la direcció C-terminal del barril).

• Les dues darreres capes de carbonis α del barril de la β-amilasa estan formades pels mateixos residus —tot i que no pels mateixos codons— amb independència del ser viu de procedència.

• Alguns dels residus de les capes 2, 3 i 4 —i en posicions molt específiques dins d'elles— són comuns per a totes les plantes i per a tots els microorganismes, però no pel conjunt de β-amilases. Aquesta diferenciació entre plantes i microorganismes indicaria que la primera β-amilasa —l'enzim ancestre— va originar una molècula per a cadascun dels dos grups evolutius. Posteriorment els membres de cadascun dels dos grups han evolucionat a partir de la corresponent β-amilasa de segona generació.

• La complexitat evolutiva que presenta la β-amilasa aconsella —com a requisit previ a qualsevol hipòtesi globalitzant— estudiar l'evolució de cadascun dels barrils TIM per separat.

• La conformació del tres darrers residus de la tira β_6 de la β-amilasa de soja —segment ^{341}FTC^{343} del motiu β_6L_6— és la mateixa si l'enzim es cristal·litza lliure o en presència de l'inhibidor α-ciclodextrina. Per la seva part, el segment presenta la mateixa conformació quan l'enzim es cristal·litza en presència dels substrats maltosa o maltal. No obstant, la conformació del segment és diferent en l'enzim lliure —o amb inhibidor— de la que presenta en els seus complexes amb la maltosa i el maltal. Això vol dir que l'unió del substrat provoca un canvi conformacional de la part C-terminal de la tira β_6. Aquesta transició comporta canvis tant a nivell de cadena principal —posició en el diagrama de Ramachandran— com de la seva cadena lateral (canvi de rotàmer de Thr_{342} i Cys_{343}). L'anàlisi comparativa de les energies potencials, les poblacions rotàmeriques i les posicions al diagrama de Ramachandran, suggereixen que la conformació corresponent al complex enzim-substrat és més estable que la de l'enzim lliure. Per tant, el mecanisme catalític de la β-amilasa, requereix —tot i que ser imprescindible per l'activitat enzimàtica— que el segment ^{341}FTC^{343} tingui una conformació metaestable per tal d'aconseguir un nivell òptim d'eficiència catalítica.

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8:
RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

Apèndix.

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8:
RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

Biblioteques de rotàmers d'alta precisió.

Deciphering the structural code for proteins: Four high-resolution rotamer libraries of amino acid residues in helices, β-strands, turns and coils.

Gerard Pujadas and Jaume Palau

Unitat de Biotecnologia Computacional, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Plaça de la Imperial Tàrraco, 1. Tarragona 43005, Catalonia (Spain).

Address for correspondence: Jaume Palau, Unitat de Biotecnologia Computacional, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Plaça de la Imperial Tàrraco, 1. Tarragona 43005, Catalonia (Spain), phone 34 (77) 55 95 65, fax 34 (77) 55 95 97, e-mail: palau@quimica.urv.es

Short title: High-resolution rotamer libraries

Subject classification: Proteins. Molecular structure.

Footnotes: a) Programs and routines (unpublished) were written in FORTRAN 77 and compiled and executed in an Indigo^2 XZ Silicon Graphics Workstation by the authors of this paper; b) rotamer libraries and comparisons of the results using the 1996 and 1994 samples may be obtained by e-mail as a set of FileMaker databases (pujadas@quimica.urv.es); c) references within the text for available unpublished rotamer libraries have a slash between the last author and the year.

Abbreviations: Brookhaven Protein Data Bank (BPDB).

Summary

In general, the quality of a rotamer library depends on the crystallographic quality of the structures used, the size of the sample and the homology threshold among these structures. Side-chain building in homology-based models, either for protein design or for folding simulations, largely depends on the quality of the rotamer library used. In our revision of the bibliography we found that the rotamer libraries described to date had not been validated, that is to say, whether libraries are significantly affected by updating protein samples or whether the data reported by different authors are equivalent or not. Moreover, for structural analysis, modelling and design purposes it also seems to be advisable to have reliable high-resolution libraries available for each element of the secondary structure (helix, β-strand, turn and coil). The purpose of the present work is to fulfil these two requirements. We demonstrate that some nonsecondary-structure dependent rotamer libraries (which we define as averaged rotamer libraries) are already stable as far as sample updates are concerned whereas other rotamer libraries included in the bibliography should not be used. We also demonstrate that published rotamer libraries derived from particular secondary-structure elements are incomplete and uncertain. Consequently, we processed four new, refined libraries specific for the different secondary structures. We validated three of these libraries (for helices, β-strands and coils). The values in our turn library are not yet stable; however, our more recent library for turns should be useful for modelling purposes. Finally, we study how rotamers in helical, β-strand and coil libraries are affected by the "syn-pentane" effect.

Keywords: rotamer library; side-chain conformation; secondary structure; coil; syn-pentane effect.

Introduction

The study of side-chain conformation in naturally occurring amino acids and the factors that govern this conformation has been an important subject of study since the middle 60's (Ramachandran & Lakshminarayanan, 1966; Ramachandran et al., 1966) and it gave rise to a series of papers published during the 60's and 70's mainly by scientists from the Centre of Advanced Study in Biophysics of the University of Madras (India). The lack of crystallised protein structures led to the first studies being made on amino acid containing compounds or short peptides as models (Ramachandran & Lakshminarayanan, 1966; Ramachandran et al., 1966; Sasisekharan & Ponnuswamy, 1970; Ponnuswamy & Sasisekharan, 1970a,b). The first studies which made use of the protein crystal structures available were made by Chandrasekaran & Ramachandran (1970) and Lipkind & Popov (1971). These authors studied the distribution of the dihedral angle values for the side-chain conformations of amino acid residues in myoglobin and tosyl-α-chymotrypsin (both studies) and lysozyme (Chandrasekaran & Ramachandran study only). Chandrasekaran & Ramachandran (1970) found that the conformations in proteins were, in general, similar to those observed in previous studies in simple amino acid and peptide structures (exceptions being the conformation of χ_2 for Asp and Asn and χ_3 for Glu and Gln). In this respect, they found that, as previously shown for amino acids and peptides (Ramachandran & Lakshminarayanan, 1966), the X_i^{γ} atom is located around 60, 180 or -60° with respect to the dihedral angle which it forms with N_i, C_i^{α} and C_i^{β} (χ_1 angle). They also demonstrated a good correlation between the experimental side-chain conformations in proteins and their theoretical distribution [obtained from a (ϕ,ψ)-plot (contact map) based on an extreme limit contact criteria for each χ_1 or χ_1/χ_2 value depending on the side-chain length]. The authors also showed that the dihedral angles of 60, 180 and -60° mentioned above are not exclusive to χ_1 and are always found in side-chain atoms bonded to a previous atom with a tetrahedral coordination. Lipkind and Popov's results show that there is good agreement between the conformations found for residues in proteins and the geometry of the optimum forms for their corresponding methylamides of N-acetyl-peptides.

At the same time, the mutual influence between the side-chain and backbone conformations was also studied (Leach et al., 1966; Kotelchuck & Scheraga, 1968; Ponnuswamy & Sasisekharan, 1970a,b; Lipkind & Popov, 1971; Sasisekharan & Ponnuswamy, 1971). The results found by Ponnuswamy &

Sasisekharan (1970a,b) suggested that the preferred backbone conformations in the N- and the C-terminal positions of a polypeptide chain partially depend on the presence and nature of side-group atoms at the beta and gamma positions (and vice versa) and not beyond. Sasisekharan & Ponnuswamy (1971) found for Phe, Tyr and Trp a correlation between the orientations of the two peptide planes and the planar group of the aromatic side-chain. Their results indicated that while the gamma atoms have definite and characteristic effects on the backbone rotational angles ϕ and ψ, the delta atoms and beyond have no effect on their preferred values. Lipkind & Popov (1971) also detected correlations between backbone and side-chain conformations in protein structures (similar to the correlations previously found for isolated dipeptides). For these authors, this demonstrates the important role that short-range interactions between backbone and side-chains play in the formation of the secondary structure of the protein. To sum up, these works revealed the existence of a mutual dependence between backbone and side-chain conformations.

Janin et al. (1978), using 19 crystallised protein structures at a resolution of 2.5Å or better, analysed the conformation of amino acid side-chains in proteins. They found that the correlation between main and side-chain conformations was only slight and they put this down to the fact that the side-chain was perturbed by steric hindrance caused by the secondary structure (and in the case of Ser and Thr, through side-chain to main-chain hydrogen bonds). Other "classical" studies of the conformations of the different side-chains were made at the end of the 70's (Bhat et al., 1979) and the first half of the 80's (Benedetti et al., 1983; James & Sielecki, 1983; Gray & Matthews, 1984).

To our knowledge, the term "rotamer" was first used in protein science as a synonym for amino acid "side-chain conformer" by Lipkind & Popov (1971), and was popularised by Ponder & Richards (1987) who associated the term with a set of residues for all amino acid types whose χ_i dihedral angles had the same side-chain spatial arrangement. In an attempt to formulate an algorithm which could supply a list of internal residue sequences compatible with a known core structure, Ponder & Richards (1987) defined the concept of "rotamer library". This consisted of a large table containing all the rotamers found for each residue and their characteristics (number of residues inside the rotamer, percentage of amino acid residues in the rotamer and mean and standard deviation for the different χ_i dihedral angles used to distribute residues in the corresponding rotamers). The characteristics of subsequent rotamer libraries (McGregor et al., 1987; Tuffery et al., 1991; Dunbrack, 1993; Dunbrack & Karplus, 1993;

Schrauber et al., 1993; Blaber et al., 1994; Dunbrack/1996 and Dunbrack/1997) in general have the same scheme. Ponder & Richards (1987) showed that 17 of the 20 amino acids (omitting Met which was very rare in their sample and Lys and Arg due to their great flexibility and poorly determined state) could be represented adequately by 67 side-chain rotamers if only χ_1 and χ_2 were used (and χ_3 for certain cases in Gln). In the same year, McGregor et al. (1987) and Piela et al. (1987) confirmed the mutual conformational constraints imposed by the backbone and the side-chain.

McGregor et al. (1987) analysed the relationship between side-chain conformation and secondary structure in globular proteins in a sample made up of 61 high-resolution (≤ 2.0Å) protein structures. Their main conclusions were: a) for most α-helical side-chains, there is no g^- conformation in the χ_1 dihedral angle distribution, and this is compensated for by a shift towards the t conformation (in comparison with non-α/non-β structures); b) in short polar side-chains inside α-helices there is a χ_1 preference for g^+ which forms hydrogen bonds with the main-chain; c) the backbone affects side-chain dihedral angles which are a long way from it (χ_2, χ_3 and χ_4). Piela et al. (1987) focused their analysis on the conformational constraints that affect the side-chains in α-helices. In agreement with results found by McGregor et al. (1987), they showed that the χ_1-rotamer with a mean value of around 60° (g^-) is nearly excluded from the α-helix by steric hindrance, whereas the relative populations of the other two rotamers, with mean values of around ±180 and -60° (t and g^+ respectively), are different from their global distribution, due to steric interactions and a reduction in hydrogen-bonding possibilities.

While searching for algorithms to quickly optimise side-chain conformations which were to be placed in a polypeptide backbone, Tuffery et al. (1991) used dynamic cluster analysis to determine the number of rotamers accessible for each amino acid. These authors worked out an averaged rotamer library, because they did not take into account the influence of secondary structure on side-chain conformations.

The increase in the number of known X-ray protein structures during the present decade has allowed more specific rotamer libraries to be developed. In an effort to determine amino acid propensities in α-helices within the context of a folded protein, Blaber et al. (1994) examined the side-chain conformation found in 100 well-refined protein structures. This study divides residues into three categories: a) residues in a nonhelical conformation; b) internal residues in helices (i.e. end

residues excluded); and c) internal residues within helices with side-chains which are at least 50% exposed to solvent (called by the authors "surface helical" residues). The corresponding rotamer libraries were produced for the nonhelical, internal helical and global samples. Their results about the negligible number of χ_1 conformations found in α-helices which can sterically hinder the hydrogen bond between i+1 and i-3 residues, coincided with those of other authors (McGregor et al., 1987; Piela et al., 1987).

Dunbrack (1993) and Dunbrack & Karplus (1993) developed backbone-dependent and backbone-independent rotamer libraries and used the former to construct protein side-chain conformations from the main-chain coordinates. Both studies demonstrated that the backbone-dependent rotamer library was considerably better at side-chain prediction than the backbone-independent library. They also made a conformational analysis of the backbone-dependent and backbone-independent rotamer preferences of amino acid side-chains based on the "butane" and "syn-pentane" effects found in hydrocarbons (Dunbrack, 1993; Dunbrack & Karplus, 1994). To study the distribution of the different χ_1-rotamers in the Ramachandran map, Dunbrack and Karplus divided it into a grid of 20 x 20° blocks and counted the number of residues in each one. Their results indicated that the χ_1-rotamer populations are unequally distributed in the Ramachandran map and that in some regions there is almost exclusively only one rotamer (although in other regions there may be more than one, well-represented χ_1-rotamer). Steric repulsions are the reason why certain conformers are scarce in certain parts of the sterically allowed Ramachandran regions.

Schrauber et al. (1993) suggested that the concept of rotamer is too strict to define real distributions of side-chain conformations in proteins. They argued that the set of amino acid side-chain conformations in globular proteins could not be considered as normally distributed around some rotamer points because outliers occur systematically. The presence of these outliers could be explained by the different environmental factors that the amino acid encounters in real protein structures. Of all the factors that influence side-chain conformation, the most important are the backbone torsion angles of the residue itself, the secondary structure in which the residue is located and the tertiary contacts with other residues far away in the sequence due to packing requirements. Despite this discrepancy, the rotamer concept has become fully accepted and is commonly used by the structural biology community.

At present, most papers published about side-chain conformations focus on their prediction and use in protein modelling [see reviews by Vásquez (1995 & 1996) and Table 1], their thermodynamic role in the protein folding process (Pickett & Sternberg, 1993; Sternberg & Chickos, 1994; Doig & Sternberg, 1995; Doig, 1996), the comparison of the statistical distribution of side-chain conformations in proteins with similar distributions calculated from the low conformational states obtained from potential functions (Nayeem & Scheraga, 1994), or the Bayesian statistical analysis of protein side-chain rotamer preferences (Dunbrack & Cohen, 1997).

One of the main aims of the present paper is to demonstrate that rotamer libraries must be validated as the high-resolution crystallised proteins increase in the BPDB. To this end we simultaneously derived two sets of rotamer libraries using the same quality cut-off and routines. The only difference between the two sets is that one was prepared from the protein structures available in March 1994 and the other from those available in November 1996. Rotamer libraries from these sets will be called Pujadas-Palau/1994 and Pujadas-Palau/1996, respectively. The quality of all libraries reported so far were successively compared with ours and we also compared both our libraries with each other and all the other libraries among one another. A second aim is to obtain and validate four reliable libraries for residues present in helices, β–strands, reverse turns and coils. The effect of "syn-pentane" interactions in the secondary-structure dependent libraries will also be discussed.

Results and Discussion

Comparative analysis of published averaged rotamer libraries. Are they exchangeable?

Most rotamer libraries available at present are averaged rotamer libraries (Ponder & Richards, 1987; McGregor et al., 1987; Tuffery et al., 1991; Dunbrack, 1993; Dunbrack & Karplus, 1993; Blaber et al., 1994; Dunbrack/1996; and, more recently, Dunbrack/1997). These libraries have been derived from different numbers of polypeptide chains [19 for Ponder & Richards (1987); 61 for McGregor et al. (1987); 53 for Tuffery et al. (1991); 132 for Dunbrack & Karplus (1993); 184 for Dunbrack (1993); 100 for Blaber et al. (1994); 393 for Dunbrack/1996; and 589 for Dunbrack/1997]. Despite this variation in sample size, two of the oldest libraries (Ponder & Richards, 1987; Tuffery et al., 1991) have been predominantly used in programs for modelling or for predicting protein side-chain conformations (see Table 1). This is surprising considering that the quality of a rotamer library should be related to the size of the sample and that the accuracy of side-chain prediction largely depends on the availability of an unbiased rotamer library (Kono & Doi, 1994).

To establish a quality control, we compared the available averaged rotamer libraries and quantified the differences. Our χ_1-averaged libraries (AV/Pujadas-Palau/1994 and AV/Pujadas-Palau/1996) were also included in these comparisons. The parameters we compared were the proportion of residues in the different conformations and their corresponding mean value and standard deviation for the dihedral angle in χ_1-based libraries. The procedures used to derive mean values and standard deviations for χ_1 from previously published data, as if χ_1 had been the only dihedral angle to be used to distribute residues into different rotamers (χ_1-based libraries), are described in Materials and Methods. Cut-off intervals and the Relative Quality Factor (RQF), both used to construct Tables 2 and 3, are also described in Materials and Methods. Differences between values for the same parameter and χ_1-rotamer in a given pair of libraries were plotted as histograms, and the value subtracted is always the one from the oldest library. Figure 1 compares all the averaged rotamer libraries available up to November 1996, from published reports and from Internet (for Dr. R.L. Dunbrack, only the 1996 library is compared), with AV/Pujadas-Palau/1996. The remaining cross comparisons, that include AV/Pujadas-Palau/1994 and the oldest Dunbrack (1993), and Dunbrack &

Karplus (1993) libraries, are not shown in Figure 1 but their most significant results are highlighted in Tables 2 and 3 respectively.

Results in Figure 1 and also in Table 2 and 3 show that not all averaged rotamer libraries are equivalent. In particular, large differences are found when libraries developed before 1993 are used in the comparison. The averaged libraries from Tuffery et al. (1991) and Ponder & Richards (1987) –widely used in side-chain modelling and prediction (see Table 1)– have considerable differences with all other libraries if comparisons are made using proportions (see Table 2), mean values (see Table 3a) or standard deviations (data only available for Ponder & Richards, see Table 3b) for χ_1. McGregor et al. (1987) yielded slightly better results with respect to χ_1 proportions than Tuffery et al. (1991) and Ponder & Richards (1987) (see Figure 1 and Table 2). It was not possible to derive mean values for χ_1 and their corresponding standard deviations from the averaged McGregor et al. (1987) library. We would like to point out the considerable differences of the $Phe_{g+,\ g-}$ proportions in the Tuffery et al. (1991) library in comparison with the rest of the libraries (see Figure 1 and Table 2). The lowest differences found are 39.6% for Phe_g^+ and 27.5% for Phe_g^- when comparisons are made with data from Dunbrack & Karplus (1993) and Ponder & Richards (1987), respectively. The χ_1 proportions for Phe_g^+ in Tuffery's results are in all cases lower than in the rest of libraries whereas the opposite is true for Phe_g^-. These differences may be due to Tuffery et al. (1991) misinterpreting the g^+ and g^- nomenclature rules for Phe (IUPAC-IUB, 1970).

Results in Figure 1 and Table 2 clearly show that χ_1 proportions reported since 1993 are almost constant, whichever author is considered. In this respect, only 17 out of 765 (2.2%) comparisons are outside the ±5% interval (see Table 2). Discrepancies are distributed among $Cys_{t,\ g}^+$, $Gln_{t,\ g}^+$, Glu_g^+ and $Trp_{g,\ g}^+$ and the highest differences found are -7.2% for Gln_t [Pujadas-Palau/1996 vs. Blaber et al. (1994)], -7.0% for Trp_g^+ [Dunbrack/1996 vs. Pujadas-Palau/1994] and -8.5% for Trp_g^- [Pujadas-Palau/1996 vs. Dunbrack/1996] respectively. Non-random differences beyond the ±5% interval for $Gln_{t,\ g}^+$ and $Trp_{g,\ g}^+$ are found in comparisons with Blaber et al. (1994) and Dunbrack/1996 libraries, respectively. These differences might be attributed to abnormal conformational distributions of the two residues within the samples used by both groups. In general, the RQFs for the 15 possible library pair comparisons have values below 2.0% (see Table 2). This indicates that the differences in the χ_1 proportions, either for the whole set of χ_1 conformations (t, g^+ and g^-) or for each one by itself, are in general low. The only exception to this general trend is

the comparison between g^+ for AV/Pujadas-Palau/1996 and Dunbrack & Karplus (1993) which has a value which is slightly higher than this cut-off limit.

Comparisons between the library pairs developed since 1993 are also good as far as the mean χ_1 value is concerned. This is reflected in Table 3a where only 11 out of 510 (2.2%) differences have values outside the ±5° interval (all of which correspond to conformations with a C_i^γ atom in the g^- position which, in general, are the least numerous in the sample; see Appendix Table 1). No differences are found in the χ_1 standard deviation beyond the ±5° interval, whether comparisons are made between Blaber et al. (1994) and Dunbrack (1993), between our two AV libraries or between Dunbrack/96 and each of our two AV libraries (see Figure 1 and Table 3b). There are significant differences between Blaber et al. (1994) and Dunbrack (1993), and the rest [14 out of 306 values (4.6%) are outside the ±10° interval]. The RQFs for the mean value and the standard deviation for χ_1 (see Table 3) are in a good agreement with the outlier results (i.e., in a comparison, the lower the RQF values are, the lower the outlier population is). It is worth pointing out that, in general, the Dunbrack/1996 standard deviations for χ_1 are lower than the equivalent values in AV/Pujadas-Palau/1996 (see Figure 1).

To sum up, if only parameters for χ_1-based libraries are considered, the quality of our two averaged rotamer libraries is similar to the latest ones available, with a few discrepancies (see Tables 2 and 3). The low frequency of outliers found in these comparisons is complemented with low RQFs whether the different χ_1 conformations are globally taken into account or not (see Tables 2 and 3).

For the record in this paper, Dunbrack's latest updated Internet library (Dunbrack/1997, bbind97.Aug.lib file, column p(r1234)) was also tested for the stability of its χ_1 parameters. These parameters were compared with those of all other χ_1-based libraries but the results were very similar to when the Dunbrack/1996 library was used (results not shown). In a recent study about the influence of backbone accuracy on conformation stability in the rotamer space (Tuffery et al., 1997) a new library (Tuffery/1997, http://duc.urbb.jussieu.fr /rotamer.html) was created. We have not made any tests on this library since the lack of information about the number of residues in each rotamer prevented us from deriving the χ_1 conformational parameters.

Validation of χ_i parameters in AV/Pujadas-Palau/1996 χ_i-based libraries

Using the same procedure as for χ_1, we compared the χ_2, χ_3 and χ_4 parameters in our AV/Pujadas-Palau/1994 and AV/Pujadas-Palau/1996 χ_i-based libraries (with "i" ranging from 2 to 4). The question arises whether we have already reached almost stable values for χ_2, χ_3 and χ_4 parameters or whether future protein sample updates may lead to a higher level of stabilization of the χ_2-, χ_3- and χ_4- rotamer libraries. As shown in Materials and Methods, χ_i was analysed by obtaining χ_i-based libraries exclusively from their corresponding χ_i dihedral angle conformations, and then calculating their χ_i parameters and comparing data obtained from the 1994 and 1996 samples. The results obtained from these comparisons showed that there are no outliers when the more restrictive interval (± 5) was considered (see Figure 2). From these results, we can conclude that the χ_2, χ_3 and χ_4 parameters in the corresponding AV χ_i-based libraries had a degree of conservation that was similar to the one for χ_1 in the AV χ_1-library. Therefore, our results demonstrate that the AV/Pujadas-Palau/1996 library has reached stable values for the χ_i proportion, mean value and standard deviation, if these parameters are analysed from AV χ_i-libraries.

Published secondary-structure dependent rotamer libraries lack of enough quality

Some authors have developed rotamer libraries from secondary-structure elements of globular proteins with several purposes: a) to analyse any existing relationship with the main-chain structure (McGregor et al., 1987); b) to determine possible deviations from rotameric χ_i values (Schrauber et al., 1993); c) to estimate variations in the side-chain entropy of an α-helix within the context of a folded protein (Blaber et al., 1994); and d) to analyse the Ncap rotamer preferences in α- and 3_{10}- helices (Doig et al., 1997). McGregor et al. (1987) observed preferred side-chain conformations inside α-helices, β-strands and non-α/non-β regions and Blaber et al. (1994) corroborated previous results in α-helices (Gray & Matthews, 1984; McGregor et al., 1987). The most important characteristic observed in α-helical rotamers was an almost complete absence of residues with a C_i^{γ} atom in the g^- position. For validation purposes, we developed χ_1-rotamer libraries (N4-C4 for helices and N3-C3 for β-strands) for the same structural elements which had previously been studied by McGregor et al. (1987) and Blaber et al. (1994) (see definitions in Materials and Methods). We did not validate the results found by Schrauber et al. (1993) or Doig et al. (1997). In the Schrauber et al. (1993) library, proportions for the

different rotamers were given only for residues which have χ_1 and, if necessary, χ_2 dihedral angle values within a ±20° interval around the corresponding canonical value. In the Doig et al. (1997) library, the number of residues in Ncap positions is not large enough to be stable with respect to sample updates. The comparative analysis was made in a similar way as for the averaged rotamer libraries and so is also based on χ_1 parameters (which have been shown to be suitable for cross comparisons when using data from different authors).

The helical parameters from McGregor et al. (1987) and Blaber et al. (1994) are significantly different, in particular for χ_1 proportions (results not shown). Figure 3a shows the validation of χ_1 parameters for the older helical libraries in comparison with the N4-C4/Pujadas-Palau/1996. From results in the figure, it appears that the Blaber et al. (1994) library is a long way from reaching stable χ_1 parameters and, therefore, we recommend that it is not used. The "α-helix centre" library by McGregor et al. (1987) is very similar to N4-C4/Pujadas-Palau/1996, as far as the χ_1 proportions are concerned. From McGregor's results, we were not able to derive individual mean and standard deviation values of χ_1 for most residues (in this case there is a grey background) and so we could not properly validate these parameters. However, in the few cases that it was possible, considerable variation was observed. By comparing N4-C4/Pujadas-Palau/1994 and N4-C4/Pujadas-Palau/1996 (Figure 3a), we concluded that the proportions, mean values and standard deviations of our newest χ_1-library are well validated, apart from a few outliers for the g^- conformation. This conformation is poorly represented in helices, as noted by McGregor et al. (1987), since there are steric clashes between the C_i^γ atom and the main-chain hydrogen bond which consists of $N\text{-}H_{i+1}$ as donor and O_{i-3} as acceptor. We also noted (results not shown) that helical residues with a C_i^γ atom in the g^- position are concentrated in the Ncap, N1 and N2 residues, and these have not been taken into account when building these helical libraries. Therefore, we conclude that stable values are reached for χ_1 parameters in the N4-C4/Pujadas-Palau/1996 rotamer library.

Similar comparisons for β-strands are shown in Figure 3b. The McGregor et al. (1987) "β-sheet centre" library and the N3-C3/Pujadas-Palau/1994 were compared with N3-C3/Pujadas-Palau/1996. In the first case, the comparison reveals large discrepancies, as well as a lack of information for most residues (in this case there is a grey background). Unlike the results shown in Figure 3a for helices, our two β-strand libraries do not reach stable values for χ_1 proportions. However, stability is reached for mean values and standard

deviations for χ_1, with a few exceptions (the most significant being the g⁻ conformation).

Summarizing the results in Figure 3, we conclude that the helix and β-strand libraries developed by McGregor et al. (1987) and Blaber et al. (1994) are not of sufficient quality and should not be used as specific secondary-structure libraries in side-chain modelling or analysis.

Secondary-structure dependent rotamer libraries. Are they necessary?

In recent years, some authors have pointed out that rotamer libraries should be used which reflect the influence of the environment on the side-chain conformational preferences. "Environmental" factors have been included in rotamer libraries by using different approaches and side-chain conformational prediction improved considerably. On the basis of earlier results about the mutual influence between the side-chain and backbone conformations (see the Introductory remarks), Dunbrack (1993) and Dunbrack & Karplus (1994) analysed the relation between main-chain and side-chain conformation on the basis of well-known steric repulsions observed in hydrocarbons. These studies allowed these authors to develop a backbone-dependent rotamer library where the Ramachandran map is divided into a grid of 20 x 20° Φ/Ψ blocks and side-chain conformational parameters are given for each one (Dunbrack, 1993; Dunbrack & Karplus, 1993). This means that their backbone-dependent rotamer library is not properly a secondary-structure dependent library, since Φ and Ψ angles for residues in coil and turn structures are found to be overlapped in typical areas for right handed α-helices and β-strands in the Ramachandran map. Chinea et al. (1995) proposed using position-specific rather than generalized rotamers in order to improve the quality of side-chain positioning when building models by homology. According to these authors, this may help to make modelling methods more robust. More recently, Ogata & Umeyama (1997) have incorporated information about the structural environment for the side-chains as vectors. These vectors have been obtained from a principal component analysis of X-ray structures and represent the variance of the main-chain atom positions around side-chains. All three approaches described in this paragraph are therefore related to main-chain conformation [although in a different manner to the approaches used by McGregor et al. (1987), Blaber et al. (1994) and the present paper, which are more closely related to secondary structures].

At present, the development of specific rotamer libraries for helices, β-strands, turns and coils is useful for correlating conformational propensities with

secondary structure. These libraries can have several different uses: a) side-chain conformational errors can be determined more accurately by analysing the presence of infrequent rotamers or large deviations of their χ_i dihedral angles from the canonical values; b) the effect of the binding process on the enzyme/substrate or enzyme/inhibitor surface can be studied and the conformational changes associated to catalysis evaluated and analysed (Pujadas & Palau, 1997); c) possible interactions in mutant proteins can be evaluated by molecular modelling; etc. In general, most of the side-chain modelling or prediction methods shown in Table 1 could be easily adapted to make use of these highly specific libraries.

Whether χ_i conformational parameters (proportions, mean values and standard deviations) inside the different types of secondary structure are stable and reliable and whether these parameters are different from the ones in the averaged library, are both issues that need careful attention.

Setting up of three sets of secondary-structure dependent χ_i-rotamer libraries.

Because data from previous reports (McGregor et al., 1987; Blaber et al., 1994) are scarce and, in general, do not appear to be of sufficient quality (see the third section of Results and Discussion), we worked out χ_i-based X-libraries (where "X" stands for "H", "B" or "C" in helix, β-strand and coil libraries, respectively). In order to validate results from these libraries, we developed two sets of rotamer secondary-structure dependent libraries (X/Pujadas-Palau/1994 and X/Pujadas-Palau/1996) which were made up of the above mentioned χ_i-based libraries which were then used to validate the corresponding χ_i parameters. The 1996 set of libraries contains present-day data (November 1996) whereas the 1994 set is taken as a reference for validating stability. As in previous comparisons (see Figure 2), the value subtracted corresponds to the 1994 library. Only a portion of protein chain structures are common to both samples and therefore the two rotamer library sets are expected to maintain a high degree of independence (see Table 4).

The various χ_i parameters (proportions, mean value and standard deviation, with "i" ranging from 1 to 4) in helices, β-strands and coils were compared for both χ_i-based sets and the results are shown in Figure 4. If the results in this Figure are compared with the ones in Figure 2 (which shows values for the comparison of the χ_i-based averaged libraries), difference values, in general, are higher in Figure 4. The fact that the samples for the three types of secondary structure are

much smaller than the global sample, explains why stability has apparently not yet fully been reached for the three structural rotamer libraries.

Even so, differences in helix, β-strand and coil χ_i-libraries are, in general, low for the percentages [for example, the biggest differences found are for the χ_1 values of $Cys_{t,\ g^+}$ in helices (-11.4 and 10.6% for t and g^+ respectively)]. In this sense, only a few differences were found beyond the ±5% interval (26 out of 276 comparisons, which is 9.4%). The outliers found are the following: a) $Cys_{t,\ g^+}$, $His_{t,\ g^+}$, Ser_{g^+} and $Tyr_{t,\ g^+}$ in helices, Gln_{g^+}, Glu_t, $His_{t,\ g^+}$ and $Trp_{t,\ g^+}$ in β-strands and $Cys_{g^+,\ g^-}$ and $Met_{t,\ g^+}$ in coils for χ_1; b) His_{-90} and $Met_{t,\ g^+}$ in coils for χ_2 and c) Met_t in helices, $Arg_{t,\ g^+}$ and $Met_{g^+,\ g^-}$ in β-strands and Met_{g^-} in coils for χ_3. Therefore these results are better than the ones obtained when comparing the χ_1 proportions in Ponder & Richards (1987), Tuffery et al. (1991) [which are the two libraries most frequently used in side-chain modelling and prediction (see Table 1)] and McGregor et al. (1987) with AV/Pujadas-Palau/1996. The values outside the ±5% interval in these validations are 29.4, 17.6 and 23.5%, respectively (see Table 2).

Differences for the mean χ_i values are also low. The outliers are the following: a) Arg_{g^-} (5.8°) and Met_{g^-} (13.8°) in helices for χ_1; b) Leu_{g^+} (8.5°) in helices, Met_{g^-} (-8.3°) and Trp_{-90} (-7.5°) in β-strands and Ile_{g^-} (-9.8°) in coils for χ_2; c) Arg_{g^-} (-5.1°) in coils for χ_3 and d) Lys_{g^+} (5.7°) in β-strands for χ_4. It is worth mentioning that the χ_i dihedral angle deviations which occur inside helices are associated with underrepresented χ_i-rotamers. For example, the proportions of all χ_1-rotamers in helices for Arg_{g^-} and Met_{g^-} are 4.1 and 2.8% respectively whereas the proportion of χ_2-rotamers in helices for Leu_{g^+} is 1.7%. The rest of χ_i dihedral angle deviations found in coils, and specially in β-strands, belong to quantitatively important χ_i-rotamers (Met_{g^-} and Trp_{-90} for χ_2 in β-strands, Arg_{g^-} for χ_3 in coils and Lys_{g^+} for χ_4 in β-strands represents a 8.9, 28.6, 22.3 and 17.8% of their χ_i-rotamers, respectively). In all cases, the above proportions refer to the 1996 sample. The standard deviations of mean χ_i values in the 1994 and 1996 libraries (Figure 4) show a high stability. Only five outliers for the ±5° interval were found: Gln_{g^-} for χ_1 in helices (-7.6°), Lys_{g^-} for χ_1 in β-strands (5.4°), Gln_{g^-} for χ_1 in coils (-6.0°), Met_{g^-} for χ_2 in helices (6.4°), Leu_{g^+} for χ_2 in coils (5.7°).

Summarizing the information shown in Figure 4, we can conclude that our particular χ_i-libraries for helices, β-strands and coils (accounting for χ_i

proportions, mean values and standard deviations) have reached a great level of stability.

The question arises whether there are advantages in using structurally specific rotamer libraries instead of using global ones, such as AV/Pujadas-Palau/1996 or Dunbrack/1997. According to Dunbrack & Karplus (1994) the main-chain dihedral angles Φ and Ψ influence the side-chain conformation largely by means of syn-pentane interactions of γ atoms with C'_{i-1} and N_i-H···O (Φ) on the one hand and O_i and N_{i+1} (Ψ) on the other. This indicates that the most characteristic discrepancies between averaged and secondary-structure dependent libraries are expected to be found for the χ_1 parameters. In accordance with this idea and using χ_1-based libraries, we compared the parameters found in AV/Pujadas-Palau/1996 and X/Pujadas-Palau/1996 (where "X" stands for "H", "B" or "C") and the results are shown in Figure 5. The value subtracted was the one from the averaged library. Considerable discrepancies were found for proportions and also for the other two parameters (mainly in helices). In this respect, the most noteworthy differences in the mean value and standard deviation for helices are found for the g^- conformation (and also for the Thr_t mean value). These discrepancies in helices are probably related with the presence of the main-chain hydrogen bond with N-H_{i+1} as donor and O_{i-3} as acceptor. This hydrogen bond produces steric clashes with C_i^γ atoms in g^- conformation (as is also the case for Thr_t) which need to deviate from the canonical +60° dihedral angle value in order to minimize the repulsions. In most cases, the presence of this type of steric interaction forces the helix to bend (unpublished observation). The moderate discrepancies for the mean value and the standard deviation in β-strands are probably due to the intramolecular hydrogen bonds connecting the β-sheets. It should be pointed out that the coil residues in Figure 5 have mean χ_1 values that are slightly higher than in averaged libraries.

We conclude that the use of secondary-structure dependent rotamer libraries is fully justified in the case of helices, β-strands and coils because, at least, χ_1-rotameric parameters can be described as stable values (see Figure 4) which are different from equivalent values in a validated averaged χ_1-rotamer library (see Figure 5).

The χ_i-turn rotamer libraries

By using the same procedure as in the section above, χ_i-turn libraries were made for 1994 and 1996 samples and the stability of the χ_i parameters was analyzed. By comparing both sets of χ_i-libraries (T/Pujadas-Palau/1994 and T/Pujadas-Palau/1996), it was observed that there are considerable differences between the parameters of some residues (see Figure 6a). The values beyond the ±10 interval are: a) for the proportions: His_g^+ and Met_g^+ for χ_1, $His_{-90,\ 90}$, Ile_t, Leu_t, Met_g^+ and Trp_{-90} for χ_2 and Met_g^+, $_g^-$ for χ_3; b) for the mean values: Phe_g^- for χ_1, Met_g^- for χ_2, Arg_g^- and $Met_{t,\ g}^+$ for χ_3 and Lys_g^+ for χ_4; c) for the standard deviation: Phe_g^- and Thr_g^+ for χ_1, Met_g^- for χ_2 and $Met_{t,\ g}^+$ for χ_3. One explanation for the high variations may be the low number of turn residues in both protein samples (1261 and 843, respectively if Ala, Gly and Pro residues are not considered). Another explanation could be the fact that there are different rotameric parameters at each position in the turn or that there are a variety of turn types within the sample. However, any analysis at the present time which divides the samples according to the positions or to the turn type would make statistical treatment impossible. In this sense, our 1996 turn library set has not yet been fully validated. However, as Figure 6b demonstrates, it is advisable to use the T/Pujadas-Palau/1996 χ_1-library for modelling purposes, instead of AV/Pujadas-Palau/1996 χ_1 parameters, if a turn has been predicted for a given segment of the protein sequence.

Rotamer libraries built up by combining dihedral angle conformations from χ_1 to χ_4 (full rotamers) and their validation

Rotamer libraries can also be obtained by taking all successive dihedral angles that define the side-chain conformation as a single entity (i.e. χ_1 for Asn, Asp, Cys, Phe, Ser, Thr, Tyr and Val; from χ_1 to χ_2 for Gln, Glu, His, Ile, Leu and Trp; from χ_1 to χ_3 for Arg and Met and from χ_1 to χ_4 for Lys). It should be noted that some χ_i dihedral angle distributions do not show well separated peaks [Asn and Asp (χ_2), Gln and Glu (χ_3) and Arg (χ_4)] and have not been considered in order to define the full side-chain conformations of these amino acids. The resulting rotamers may be considered as "full rotamers" in contrast with the "χ_i-rotamers" used up to now. In this way, we prepared two library sets for helical, β-strand, coil and averaged 1994 and 1996 samples. The results for the 1996 sample are presented in the Appendix (in the Appendix tables, Lys rotamers have been obtained using only dihedral angles from χ_1 to χ_3). We used a similar analytical procedure for validation as in previous sets of χ_i-libraries,

the only difference being that the stability of the mean and standard deviation values was tested for all χ_i dihedral angles used to define the corresponding rotamer. We only compared results from rotamers that were present in both the 1996 and 1994 equivalent libraries. The most outstanding features of these comparisons are discussed in the paragraphs below.

Although samples have now been split into smaller parts than when the χ_i-based libraries were derived, only a small number of rotamers had differences in their proportions outside the ±5% interval (and no outliers are beyond the ±10% interval if rotamers defined by more than one χ_i dihedral angle are considered). The outliers, classified according to the absolute values of the differences in the proportions, are: a) in the averaged sample: Cys_t (-5.2%); b) in helices: Cys_t (-11.4%), Cys_g^+ (10.6%), Tyr_g^+ (-6.4%), Ser_g^+ (5.9%), $Glu_g^+{}_t$ (-5.7%), Tyr_t (5.3%), and $Arg_g^+{}_{tt}$ (5.1%); c) in β-strands: Gln_{tt} (-7.8%), $Met_g^+{}_{tg}^-$ (6.2%), $Gln_g^+{}_t$ (5.9%), $His_g^+{}_{90}$ (-5.8%) and His_{t90} (5.7%); d) in coils: Cys_g^- (7.9%), Cys_g^+ (-6.7%), $Met_g^+{}_g^+{}_g^+$ (6.4%), $Trp_g^+{}_0$ (-5.9%), $Trp_g^+{}_{90}$ (5.7%), $Gln_g^-{}_t$ (-5.3%) and $His_g^+{}_{90}$ (5.1%).

Other characteristics that deserve some comment as far as the mean values are concerned are: a) an appreciable number of rotamers show differences that are outside the ±5° interval, for at least one mean χ_i value, (74 out of 198 total rotamers in the averaged sample, 72 out of 170 in helices, 66 out of 158 in β-strands and 79 out of 172 in coils); b) deviations found in helices are not apparently related to the steric hindrance caused by the main-chain hydrogen bond between $N\text{-}H_{i+1}$ and O_{i-3} [only a third of the 39 outliers found for χ_1 are rotamers with a C_i^γ atom in g^- position]; c) only a few of the outliers for the ±10° interval are quantitatively important [Trp_{t0} (5.4%) in the global sample; Met_{ttg}^+ (4.9%) in helices; Trp_{t0} (6.6%) and $Lys_g^+{}_{ttg}^-$ (5.2%) in β-strands; and Trp_{t0} (5.7%) and $Met_g^-{}_{tg}^+$ (5.3%) in coils; the number in parentheses indicates the proportion in the corresponding 1996 library]; d) an appreciable number of rotamers had stable mean values for all their χ_i dihedral angles (all deviations within the ±5° interval) and most of these rotamers contain more than 5% of the residues for the corresponding amino acid in the 1996 samples (73 out of 198 total rotamers in the global sample, 58 out of 170 in helices, 71 out of 158 in β-strands and 74 out of 172 in coils). As far as the standard deviation of the rotamer dihedral angles are concerned the differences are, in general, larger than those for proportions and mean values. Moreover, some of the outliers for the ±5° interval are in quantitatively important rotamers: a) $Met_{tg}^-{}_g^-$ (5.8%) for the averaged sample; b) $His_{t\text{-}90}$ (15.4%) for helices; c) $His_{t\text{-}90}$ (15.6%), $Met_g^+{}_{tg}^-$ and

Met_{ttg}^{+} (11.9%), $Gln_{g}^{+}g^{+}$ (11.6%) and Trp_{t90} (10.2%) for β-strands; and d) $Met_{g}^{+}tg^{-}$ (18.6%), Glu_{tt} (13.5%), $Met_{g}^{+}g^{+}g^{+}$ (12.4%) for coils.

Disregarding turns, it can be concluded that the 1996 full rotamer libraries have reached a stability which is similar to that of χ_i-based libraries. Therefore, the 1996 specialised libraries (for helices, β-strands and coils) should be suitable –better than older libraries– for modelling and for predicting side-chains in particular secondary structures.

The "syn-pentane" effect in helical, β-strand and coil rotamer libraries

Dunbrack (1993) and Dunbrack & Karplus (1994) demonstrated that the mutual dependence between main-chain and side-chain conformations may be satisfactorily described by well-known steric repulsions observed in hydrocarbons: the "butane" and the "syn-pentane" effects. They argued that the "syn-pentane" effect is not only restricted to the Φ/χ_1 and Ψ/χ_1 dependence but that it may also extend along the whole side-chain and, therefore, strongly determine which $\{\chi_i, \chi_{i+1}\}$ dihedral angle combinations are, or are not, allowed.

The "syn-pentane" effect occurs only when there are two consecutive dihedral angles in gauche conformation, one of which is g^+ and the other g^-; so we investigated the quantitative importance in helices, β-strands and coils for the $\{\chi_1, \chi_2\}$ combinations which may cause this type of steric repulsion. It is worth pointing out that Dunbrack & Karplus (1994), and more recently Dunbrack & Cohen (1997), have studied the occurrence of $\{\chi_i, \chi_{i+1}\}$ pairs affected by "syn-pentane" in their corresponding averaged libraries. The dihedral angle pairs which we have studied are coincident with the ones used in these papers. In the following results, the dihedral angle conformations are expressed as a function of χ_1 and χ_2 so that they can be more easily identified in the full rotamer libraries (see Appendix tables). The $\{\chi_1, \chi_2\}$ combinations are the following: a) $\{g^-, g^+\}$, $\{g^+, g^-\}$, $\{t, g^+\}$, $\{g^-, g^-\}$ for Arg, Gln, Glu, Ile, Leu ($C_i^{\delta 1}$), Lys and Met; b) $\{g^-, t\}$, $\{g^+, g^+\}$, $\{t, t\}$, $\{g^-, g^+\}$ for Leu ($C_i^{\delta 2}$). It should be pointed out that there are "syn-pentane" interactions produced by $\{\chi_1, \chi_2\}$ pairs other than $\{g^-, g^+\}$ or $\{g^+, g^-\}$ when atoms which are not used in the definition of χ_1 or χ_2 (such as C'_i and $C_i^{\delta 2}$) participate in the steric repulsions. For the $\{\chi_2, \chi_3\}$ pairs for Arg, Met and Lys, and $\{\chi_3, \chi_4\}$ for Lys (not studied in this paper), the combinations are $\{g^-, g^+\}$ and $\{g^+, g^-\}$.

In the following results, we have quantified the importance of these $\{\chi_1, \chi_2\}$ pairs as the proportion of residues with the same χ_1 conformation that belong to

a specific χ_2 conformation. Therefore, these proportions are correlated with the conditional probability $p(\chi_2|\chi_1)$ and are analogous to the concept of conditional backbone-independent rotamer developed by Dunbrack & Cohen (1997). The values in parentheses and after the percentages, indicate the number of residues for the $\{\chi_1, \chi_2\}$ pair. The most noteworthy characteristics are the following: a) $\{g^-, g^+\}$ is only quantitatively important for Glu in helices and coils [41.7% (30) and 37.6% (44) of Glu residues]; b) the proportions for $\{g^+, g^-\}$ and $\{t,g^+\}$ are low [the maximum values found are 17.2% (90) for $Leu_g{}^+{}_g{}^-$ in β-strands and 8.8% (17) for $Lys_{tg}{}^+$ in coils]; c) $\{g^-, g^-\}$ is only relatively important for Leu in coils and β-strands [52.6 (10) and 51.7% (15), respectively] but it is less than 15% for the rest (with a maximum number of 7 residues for Ile in coils); d) the $\{g^-, g^+\}$ pair does not appear in the Leu sample and the same is true of the Leu $\{g^-, t\}$ pair in helices (see Appendix Table 1); e) the few Leu residues that have a g^- conformation for χ_1 in β-strands and coils are distributed in the $\{g^-, t\}$ and $\{g^-, g^-\}$ pairs (14 and 15 residues, for β-strands; and 6 and 10 residues for coils, respectively); f) when χ_1 for Leu is in the g^+ conformation, then the $\{g^+, g^+\}$ is very rare (18 out of 1,006 residues in helices, 16 out of 524 in β-strands and 13 out of 766 in coils); g) if χ_1 for Leu is in the t conformation, then the preferred conformation for χ_2 is g^- which is the only one in which "syn-pentane" interactions do not occur (if χ_2 is g^+, the atoms responsible for the "syn-pentane" effect are C'_i and $C_i^{\delta 1}$ whereas if the conformation is t, then the effect is produced by the interaction of C'_i with $C_i^{\delta 2}$). This is confirmed by our experimental results which showed that the range of proportions for Leu $\{t,t\}$, $\{t,g^+\}$ and $\{t,g^-\}$ are 11.8–8.3, 2.5–0.6 and 90.6–84.2% respectively (the percentages being taken from data in helices, β-strands and coils). In conclusion, the $\{\chi_1, \chi_2\}$ pairs which are expected to be influenced by the "syn-pentane" effect are not quantitatively important whatever their secondary-structure origin (see Appendix Table 1).

If we compare our results for the $\{\chi_1, \chi_2\}$ pairs affected by the "syn-pentane" effect with Dunbrack & Karplus' (1994), Dunbrack & Cohen's (1997) and Dunbrack/1997's (bbind97.Aug.condlib file obtained from Dunbrack's website, see the corresponding Materials and Methods section) the most significant conclusions are: a) the range of proportions for rotamers affected by this type of steric interaction in helices, β-strands and coils (from 4.7 to 2.6% for Arg, 10.0 to 7.2% for Gln, 18.1 to 3.5% for Glu, 3.6 to 1.3% for Ile, 19.9 to 12.1% for Leu, 6.3 to 3.6% for Lys and 7.3 to 2.3% for Met) are in good agreement with the results described by Dunbrack & Karplus (5% for Arg, 9% for Gln, 13% for Glu, 5% for Ile, 18% for Leu, 7% for Lys and 6% for Met) and Dunbrack/1997

(4.6% for Arg, 8.3% for Gln, 12.7% for Glu, 4.1% for Ile, 17.0% for Leu, 6.2% for Lys and 4.5% for Met); b) the low number of g^- χ_1-rotamers for Leu in helices (0.2%), β-strands (2.9%) and coils (1.9%) is also consistent with the results obtained by Dunbrack & Karplus (1994) and Dunbrack/1997, who found that only less than 2% of rotamers for Leu have their χ_1 angle within the +60±60° interval; c) deviations from the canonical dihedral angle values are found in helices, β-strands and coils (helices: from -41.2 to 44.5° and from -47.6 to 47.6°; β-strands: from -34.1 to 17.3° and from -39.3 to 32.8°; coils: from -40.3 to 45.4° and from -31.0 to 39.8°, for χ_1 and χ_2 ranges respectively) although in a wider range than the 10-30° interval described by Dunbrack & Karplus (1994) or the range obtained from Dunbrack/1997 data (from -32.6 to 24.5° and from -35.9 to 33.3°, for χ_1 and χ_2 ranges); d) for Phe and Tyr, our results when χ_1 has a t or a g^- conformation for helices, β-strands and coils are similar to those obtained for Phe by Dunbrack & Karplus (1994), Dunbrack/1997 and Dunbrack & Cohen (1997) in their averaged libraries (the mean χ_2 values for Phe and Tyr when χ_1 has a t conformation are: 76.4 and 77.6° for helices, 75.1 and 72.9° for β-strands, 73.7 and 74.3° for coils; the mean χ_2 values for Phe and Tyr when χ_1 has a g^- conformation are: 85.0 and 84.6° for helices, 92.7 and 91.0° for β-strands, 89.0 and 89.4° for coils); e) when the χ_1 dihedral angle for Phe and Tyr is in the g^+ conformation, our results show a strong discrepancy between mean χ_2 values for helices and coils (106.4 and 113.2° in helices, and 103.3 and 103.1° in coils for Phe and Tyr respectively) and for β-strands (94.4 and 93.7° for Phe and Tyr respectively). With respect to this last point, it is worth mentioning that values from the Dunbrack & Karplus (1994), Dunbrack/1997 or Dunbrack & Cohen (1997) averaged libraries for Phe (99, 97.9 and 100°, respectively) are between our mean χ_2 values for helices and coils on the one hand and for β-strands on the other (and therefore have the expected behaviour for an averaged sample). The same is true for Tyr if Dunbrack/1997 data are used. We have not found any plausible interpretation for this discrepancy because χ_2 is not apparently affected by main-chain conformation.

In conclusion the results presented above, for both percentages and for mean dihedral angle values, are in general only slightly affected by the secondary-structure origin of the sample. So few discrepancies may be due to sample size limitations and therefore we conclude that, as previously demonstrated for averaged libraries (Dunbrack & Karplus, 1994), the "syn-pentane" $\{\chi_1, \chi_2\}$ interactions are not only independent of the backbone dihedral angles but also of the secondary-structure environment. Other points of interest such as the

comparison of the χ_2, χ_3 and χ_4 distributions in helices, β-strands and coils (which are expected to be nearly equal if the "syn-pentane" interaction is the long-range force that drives the side-chain conformation) or the analysis of the experimental results in our different rotamer libraries for $\{\chi_2, \chi_3\}$ and $\{\chi_3, \chi_4\}$ pairs affected by the "syn-pentane" effect and their comparison with Dunbrack/1997 $p(\chi_2|\chi_3)$ and $p(\chi_3|\chi_4)$ data will be the subject of further experimental attention in our research group.

Final remarks

In this paper we demonstrate that widely-used, averaged rotamer libraries (McGregor et al., 1987; Ponder & Richards, 1987; Tuffery et al., 1991) have clear inconsistencies in some of their χ_1 parameters. Such dissimilarities have been overcome by the averaged libraries developed since 1993, which in most cases show comparable values for equivalent χ_1 parameters. This is a good example of knowledge-based research which yields high quality libraries with stable parameters, if the sample size is large enough. By increasing the sample size the parameters do not change anymore.

To our knowledge, incomplete libraries for residues located in helical and β-strand regions have only been reported by McGregor et al. (1987), Schrauber et al. (1993) and Blaber et al. (1994). The scarcity of results and the lack of validation for McGregor and Blaber libraries fully justify the development of new secondary-structure dependent libraries (adding turn and coil libraries), by increasing the protein sample, using new refined programs and algorithms and studying the full secondary-structure length. With the exception of the turn library, our secondary structure libraries for full rotamers have stable χ_i parameters, as does our averaged library. It should be noted that the χ_1 parameters for the specific libraries clearly diverge from the averaged one, which shows that the latter is a linear combination of the secondary-structure dependent libraries.

The increasing number of structures deposited in the BPDB will lead to more specific studies. One example is the analysis of the influence of the helical main-chain hydrogen bond on rotamers and the subsequent development of different helical libraries as a function of this influence (Ncap, N1-N2, N3, internal helical, C3-C1 and Ccap libraries). In this line of research, Doig et al. (1997) recently studied the N-termini side-chain rotamer preferences in helices; they reported strong energetic and structural preferences for the N-termini rotamers which are very different from the results for positions within helix

interiors. Unpublished results in our group for Ncap, N1-N2, N3, internal helical, C3-C1 and Ccap helical segments also suggest strong and differentiated rotamer preferences (Pujadas & Palau, in preparation).

Finally we would like to point out the importance of quantitative validation before accepting knowledge-based results. Studies and conclusions which are not validated in this way, although important and necessary from a qualitative point of view, must be considered carefully because they may be conditioned by sample limitations and therefore vary in future sample updates. At present, we recommend our specialised full rotamer libraries for prediction and modelling purposes.

Materials and Methods

Selecting the initial unrefined sample

The latest 1996 release of the PDB_Select (pdb_select.1996-Nov-11 file) (Hobohm et al., 1992; Hobohm & Sander, 1994) was used as a source to prepare a working sub-bank of subunit structures. The file was taken from the EMBL ftp anonymous server via Internet (ftp address: ftp.embl-heidelberg.de; directory location: /pub/databases/pdb_select). To ensure the statistical representativity of the results obtained from current analytical approaches, the maximum homology cut-off allowed between structures, in the PDB_Select file, was 35% (which provides 842 subunits). From this set, only structures with X-ray diffraction high quality parameters (resolution $\leq$ 2.0Å and R-factor $\leq$ 0.200) were considered. Subunits lacking side-chain coordinates were also removed from the sample (2 subunits). In this way, the first working sample to be purified was a set of 314 protein subunits belonging to 304 different BPDB files.

Refining the initial sample

The presence of 35 subunits in the working sub-bank with nondefined secondary-structure limits was revealed using our **LSSL** routine (**L**ook for **S**econdary **S**tructure **L**imits). In these cases, the limits were obtained by using the DSSP algorithm (Kabsch & Sander, 1983) applied to the corresponding BPDB files. The information for the secondary-structure limits was taken from each DSSP file using our **JR_BPDB_DSSP** routine (**J**oin **R**elated **BPDB** and **DSSP** files) and incorporated into the corresponding BPDB document with a style that matched the Brookhaven format. Therefore, after the application of

the **JR_BPDB_DSSP** routine, all 314 subunits afforded information about secondary-structure limits in a common format and were ready for use in the programs that follow.

The coordinates of the working subunit sample were carefully examined with our **REF_COORD** routine (**REF**ine **COORD**inates) in order to prevent the presence of "irregularities". More specifically, the coordinates eliminated due to "irregularities" belong to: a) nonstandard protein amino acids; b) residues which lack at least one of their atoms; c) residues with more atoms than the standard number (due in most cases to chemical modifications); d) residues with all their atoms but not ordered in the standard mode; e) heteroatom coordinates (water, substrates, inhibitors, cofactors, etc.). Once the coordinates had been "purified", our routine **SEP_COORD_ST** (**SEP**arate **COORD**inates from **ST**ructures) was used to distribute the coordinates of helices, β-strands, turns and coils into separate files according to the secondary-structure type each one belongs to. Secondary structures labelled HELIX in the BPDB files were pooled together; therefore, no distinction was made concerning the type of the helix (α or 3_{10}). The same criterion was applied for structures labelled TURN in the BPDB. Before being accepted, the secondary structures were subjected to an exhaustive quality control by the self **SEP_COORD_ST** routine. Secondary structures with the following errors were rejected: a) incomplete secondary structures (due either to the lack of coordinates for at least one of their residues or to the lack of coordinates in the preceding or following residues which are necessary if the Φ and Ψ dihedral angles for the first and last residue of the secondary structure, respectively are to be calculated); b) secondary structures superimposed on other secondary structures of the same or different type; c) β-strands and helices which are less than three and five residues long respectively; d) turns which are not four residues long. Once the quality test had been made, the number of subunits which provided at least one secondary structure to the sample coordinate files was 262 (see Table 4). This was the definitive database used in our study. Of these, 222, 218, 105 and 258 subunits contribute at least with one helix, one β-strand, one turn or one coil, respectively. The characteristics of the study sample are: a) number of secondary structures: 1,262 helices, 1,769 β-strands, 450 turns and 2,959 coils; b) range in length of secondary structures other than turns: 5–48 residues for helices, 3–22 for β-strands, 2–44 for coils, and c) total number of sample residues: 43,698. The residue distribution in each secondary structure is shown in Table 5.

When necessary, the coordinates were corrected to comply with rule 2.3.2 of the IUPAC-IUB Commission (IUPAC-IUB, 1970) which concerns the atomic labels of the symmetric aromatic rings (Phe and Tyr). Corrections were made with the **CSARL** program (**C**orrect **S**ymmetric **A**romatic **R**ing **L**abels). A number of 884 Phe [48.3% of all Phe residues, of which 283 were in helices (43.2%), 262 in β-strands (48.0%), 28 in turns (58.3%) and 311 in coils (53.6%)] and 821 Tyr [48.8% of all Tyr residues, of which 253 were in helices (47.0%), 238 in β-strands (43.3%), 35 in turns (66.0%) and 295 in coils (54.3%)] respectively, were corrected (the proportions in brackets indicate the percentages of Phe or Tyr residues that were corrected for each secondary structure). Val and Leu residues, which are known to be susceptible to non standard labelling of the γ and δ carbons, respectively, were found to be correctly labelled in all cases (results not shown) and therefore no correction was made.

The corrected coordinates for residues were deposited in different files according to their amino acid and secondary-structure origin. This split process was done with the **SRC_SSF** (**S**plit **R**esidue **C**oordinates from **S**econdary **S**tructure **F**iles) routine. As a result, eighty sets of coordinates (20 for each secondary structure, 4 for each amino acid) were obtained. Twenty merged documents were prepared by pooling the four coordinate files from the same amino acid, using the **POOL_RES** (**POOL RES**idues) routine. The pooled files are useful for comparing results obtained from residues that belong to specific secondary structures with those obtained from a global nonsecondary-structure dependent sample.

Building the rotamer libraries

From the previously pooled coordinate files, we used the program **CAL_DH_AN** (**CAL**culate **DiH**edral **AN**gles) to draw graphs showing the distribution of the values for the χ_i dihedral angles needed to assign residues to the different rotamers. The routine calculates the χ_i values by following the IUPAC-IUB rules (IUPAC-IUB, 1970). Limits for the different χ_i conformations were established for each residue by visually inspecting the graphical distribution of their values, without considering the secondary structure of origin (results not shown). In order to simplify our conformational analysis of the different side-chains, we gave the concept of "rotamer" a flexible definition based on the χ_i dihedral angles which are used to distribute residues into the different rotamers: a) when the full set of dihedral angles which make

up the side-chain conformation of each amino acid (from χ_1 to χ_4 when necessary) are used, then the rotamers are considered to be "full rotamers"; and b) when only one of the χ_i dihedral angles is used (thus making it easier to study the conformation of each χ_i angle, independently from the rest of χ_i), then the rotamers are considered to be "χ_i-rotamers". For some residues (Asn, Asp, Cys, Phe, Ser, Thr, Tyr and Val), χ_1- and full rotamer definitions coincide. In addition, χ_2 data for Asn and Asp and χ_3 data for Gln and Glu were not used to build rotamers because their dihedral angle distributions do not have well separated peaks.

Residue coordinates were distributed among their corresponding rotamers by applying the program **DTB_RES_ROT** (**DisTriB**ute **RES**idues in their **ROT**amers). The algorithm this program uses has a two step procedure. In the first step, the complete set of coordinates for a residue is extracted and all the χ_i dihedral angles which have previously been selected as defining either the χ_i- or the full rotamer concept are calculated. In the second step, the coordinates are assigned to a rotamer by checking which χ_i conformation each χ_i dihedral angle in the current rotamer concept definition belongs to. The **DTB_RES_ROT** routine was applied to the coordinate files obtained by the **SRC_SSF** and **POOL_RES** programs (with the exception of the files containing Ala, Gly and Pro). As a result of these processes, we obtained a collection of documents which contained the full set of coordinates from residues that belong to the same secondary structure, amino acid and rotamer (or only the same amino acid and rotamer if pooled files were used). These documents constituted our rotamer data bank. The corresponding rotamer data bank was calculated for each rotamer definition needed.

The different rotamer data banks were used as a source of statistical information which was extracted with the program **STAT_ROT** (**STAT**istics of **ROT**amers). The program calculates several characteristics for each rotamer of the different amino acids: a) the total number of residues and the percentage this represents of the whole set of residues from the same amino acid (whatever its structural origin may be); b) the number and proportion of residues within each secondary structure; c) how residues in the rotamer are distributed among helices, β-strands, turns and coils; d) the mean value, standard deviation, and the minimum and maximum experimental values for the whole set of dihedral angles used to define the rotamer concept for the corresponding amino acid in the rotamer data bank (these dihedral angle parameters are given for the global, helical, β-strand, turn and coil residues in the rotamer). All the resulting data

were exported to a Macintosh database file (FileMaker Pro v2.1; Claris Corporation) for ease of handling.

Comparison between pairs of rotamer libraries

One technical object of our work was to compare all possible pairs of rotamer libraries presently available in order to determine differences between them and, therefore, the relative quality of one and the other. The libraries used for comparisons were obtained from the bibliography (McGregor et al., 1987; Ponder & Richards, 1987; Tuffery et al., 1991; Dunbrack, 1993; Dunbrack & Karplus, 1993; Blaber et al., 1994) and from Internet (Dunbrack/1996 and Dunbrack/1997; http://www.cmpharm.ucsf.edu/~dunbrack/; bbind96.Jan.rawlib and bbind97.Aug.lib files, respectively). They were all compared with our present results (Pujadas-Palau/1996). Comparisons, when possible, were made only for the χ_1 dihedral angle parameters (proportions, mean value and standard deviations) derived from data found in the original references mentioned above. This derivation was carried out by joining, in each case, rotamer residues with the same χ_1 dihedral angle conformation.

Current statistical sampling expressions were used to calculate the mean values and standard deviations for χ_1 after pooling. The equations, shown below, transform the χ_1 statistical parameters reported for rotamers whose χ_1 conformation is the same but whose whole side-chain conformation is different, into χ_1 parameters for the pooled sample. These parameters can then be compared with the ones in our χ_1 libraries (those built from χ_1-rotamers).

$$\bar{x} = \sum_{i=1}^{r} \frac{N_i}{N} * \bar{x}_i \qquad (1)$$

$$s^2 = \frac{\sum_{i=1}^{r} \sum_{j=1}^{N_i} x_{ij}^2 - N * \bar{x}^2}{N-1} \qquad (2)$$

where (1) defines the mean value for χ_1 and (2) its variance. In the case of equation (2), the value for $\sum_{j=1}^{N_i} x_{ij}^2$ was found using the following expression:

$$\sum_{j=1}^{N_i} x_{ij}^2 = s_i^2 * (N_i - 1) + N_i * \bar{x}_i^2 \qquad (3)$$

The notation used was: r: the number of initial rotamers which were to be pooled; N_i: the number of residues in the rotamer i; N: the number of residues after pooling; $\bar{x}_i$: mean χ_1 value in rotamer i; x_{ij}: χ_1 values for each residue in rotamer i; s_i: standard deviation for the χ_1 value in rotamer i.

When the number of residues in rotamers was not indicated by the authors, but compositional percentages were available (as in the case of Tuffery et al., 1991), the equation used instead of (1) was:

$$\bar{x} = \sum_{i=1}^{r} \left(\frac{P_i}{\sum_{i=1}^{r} P_i} * \bar{x}_i \right) \qquad (4)$$

where P_i is the percentage of residues in rotamer i for a given amino acid.

In order to compare helical and β-strands results from our present polypeptide sample (262 subunits) with previously published results (McGregor et al., 1987; Blaber et al., 1994), we created segmental helix and β-strand rotamer libraries by using criteria and definitions from these authors (and using the same working protocol and programs described above to obtain our present rotamer libraries). Although labeled with different names ["α-helix centre" by McGregor et al. (1987) and "internal helical" by Blaber et al. (1994)], the helical segments considered by both groups are equivalent. Only McGregor et al. (1987) developed a library for β-strand segments, labelled "β-sheet centre". In this way, we worked out χ_1-rotamer libraries for β-strands and helices (which accounts for residues that are at least, three positions away from β-strand ends and four positions away from helix ends) and the χ_1 parameters were compared with equivalent data from McGregor et al. (1987) and Blaber et al. (1994) obtained using the above expressions (1-3). These libraries were named N3-C3 (β-strands) and N4-C4 (helices) depending on the secondary-structure segment from which they were derived.

In all cases, comparisons between pairs of libraries were made by subtracting parameter values (proportion, mean χ_1 value and standard deviation) for the same rotamer in one library from its equivalent ones in another. The difference values were always obtained by subtracting the oldest from the newest data. To evaluate the similarity of a pair of rotamer libraries we have chosen two different measures. The first of these is the number of differences that are beyond an arbitrary cut-off interval. It provides specific information about which conformations have the least similar values of the two libraries compared. The arbitrary intervals chosen were ±5, ±10 and ±15 for different deviation levels. The second measure is the average of the absolute values for the differences. This may be calculated globally from the whole set of χ_1 conformations (t, g^+ and g^-) or for each one by itself. We called this parameter the "Relative Quality Factor" (RQF). Both measures were shown to be very

useful when determining the relative similarity between a pair of libraries because they give complementary information. In this respect, the former enumerates the outliers, if any, whereas the latter quantifies the overall similarity between a pair libraries.

Variability of the rotamer library as a function of updating the protein working set

The dependence of the rotamer library results on updating the protein working set was investigated by comparing the present results with the ones obtained using a nearly three year old PDB_Select file (pdb_select.1994-mar). The quality criteria, parameters and programs used to obtain both sets of rotamer libraries (1994 and 1996) were exactly the same. The subunits of entries that are obsolete in the current BPDB release were replaced by the new entries to avoid possible conformational mistakes which have now been corrected (1ABK by 2ABK, 1AYH by 2AYH, 1END by 2END, 1GLT by 2GLT and 1SIM by 2SIM). The number of subunits in the 1994 sample which provided at least one secondary structure to the sample coordinate files was 134 (see Table 4). Of these, 109 contribute with at least one helix, 108 with at least one β-strand, 69 with at least one turn and 133 with at least one coil. The characteristics of the sample are: a) number of secondary structures: 574 helices, 778 β-strands, 298 turns and 1,331 coils; b) the length ranges from: 5–33 residues for helices, 3–15 for β-strands and 2–36 for coils (the length of all turns is, as previously noted, 4 residues) and c) total number of sample residues: 19,454 (6,719 in helices, 4,642 in β-strands, 1,192 in turns and 6,901 in coils; see Table 5). There are 81 identical structure subunits used in both libraries (see Table 4). Of the rest: a) 28 are subunit codes which have been replaced in the 1996 PDB_Select file by equivalent entries (same enzyme from the same or different source) with usually better structural quality; and b) 154 are subunits in the 1996 sample with no equivalent partner in the 1994 sample (whereas the opposite is true for 25 subunits in the 1994 sample). A comparison between the residue distribution and the secondary-structure distribution (see Table 5), and the number of secondary structures of each type in the two rotamer library sets shows that the 1996 sample is approximately twice as big as the 1994 sample. It is also worth mentioning that the maximum difference between the percentages of each amino acid within the different secondary structures and within the global sample, when comparing results from 1996 and 1994, is as low as 1.1% (for Ser in turns, see Table 5). All rotamer libraries (AV, H, B, T, C, N3-C3 and N4-C4) in the 1994 and 1996 sets were worked out in the same way, which enabled

comparisons to be used for validation. The methodology used to evaluate the differences between our 1996 and 1994 library sets was the same as the one described in the previous section but, in this case, the comparisons were extended to the different χ_i- and full rotamer libraries.

Acknowledgements

We thank Prof. Enric Querol for computational facilities, Drs. Jordi Grifoll and Teresa Segués for a thorough discussion concerning statistical aspects of the paper, Mrs. Eva Sugrañes for help in compiling rotamer libraries available in the references, and John Bates (English Text & Style Correction Service from our University) for his help during the manuscript writing process. G.P. has been the recipient of a fellowship (FI/93-129) from the Catalan Governmental Agency CIRIT (Generalitat de Catalunya). We thank general facilities from our University, although this work has not been awarded grants by any research-supporting institution.

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8:
RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

Figures de l'apèndix.

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8:
RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

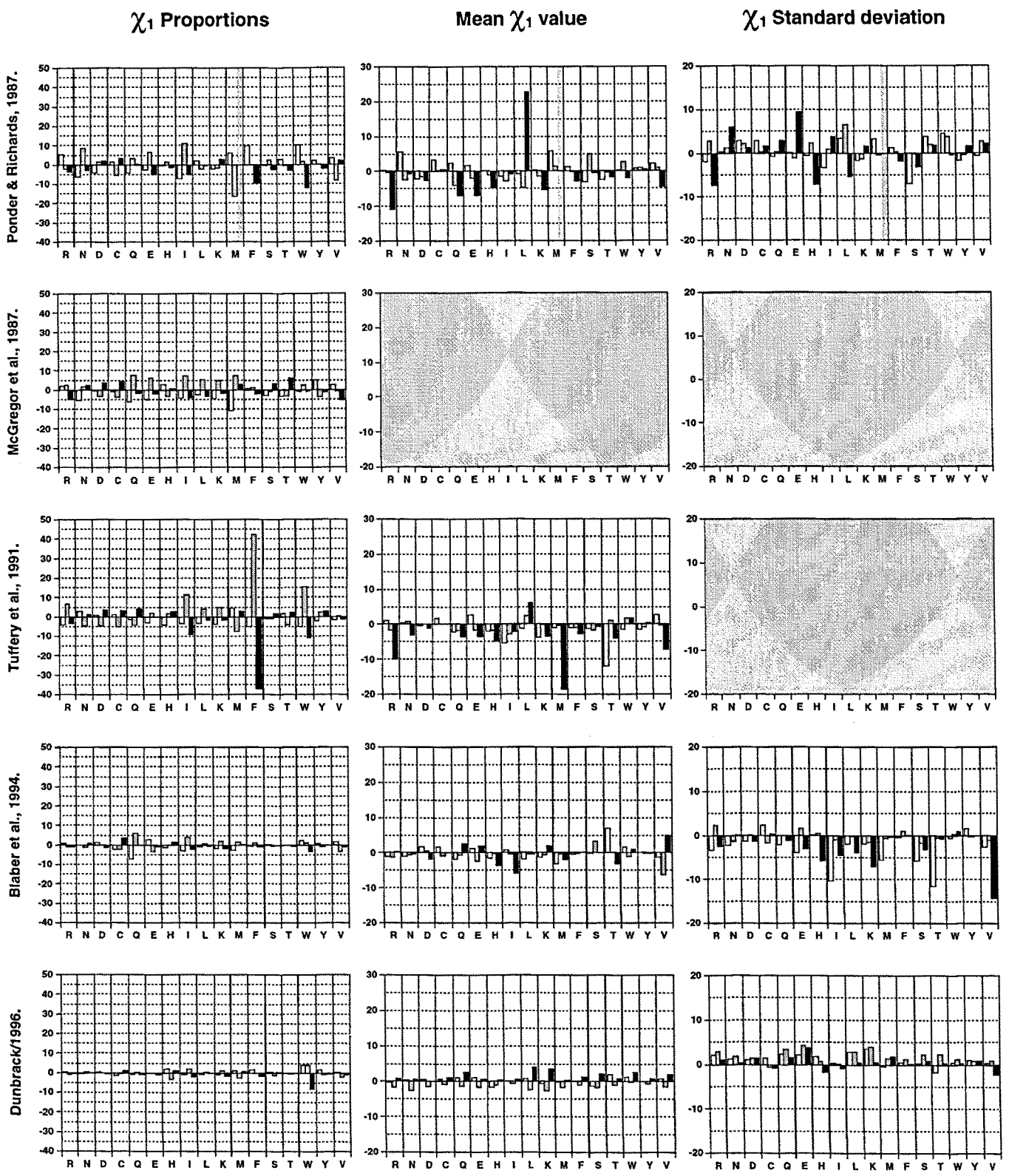


Figure 1. Comparisons between the χ_1 parameters (proportion, mean value and standard deviation) in the AV/Pujadas-Palau/1996 library and the averaged rotamer libraries available up to November 1996, from published reports and from Internet (for Dr. R.L. Dunbrack, only the 1996 library is compared). All data used in these comparisons are from χ_1-based rotamer libraries. The name of the library compared with our data is shown in the left hand margin of each row. The values shown were obtained by subtracting data found in any particular library from equivalent values in AV/Pujadas-Palau/1996. The white, grey and black columns indicate differences for t, g^+ and g^- χ_1 conformations respectively. A grey background in the plots indicates that there is no equivalent data in the library that is compared with our results.

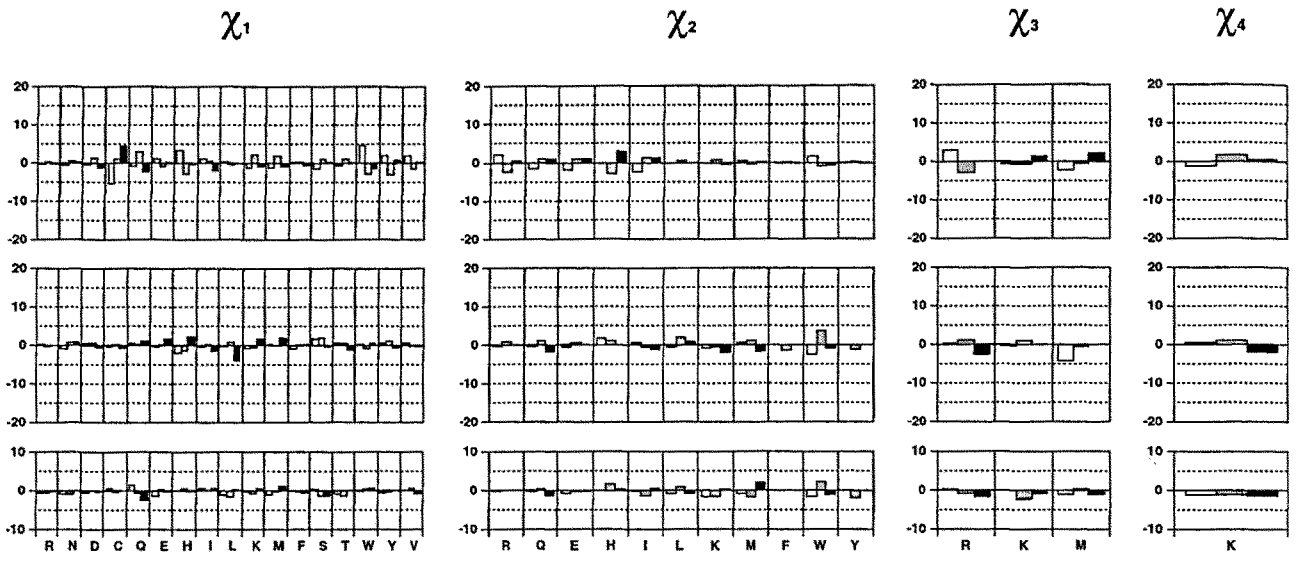


Figure 2. From top to bottom, comparisons between the χ_i parameters (proportion, mean value and standard deviation) in the AV/Pujadas-Palau/1996 and the AV/Pujadas-Palau/1994 χ_i-based rotamer libraries. The values shown are the differences between 1996 and 1994 equivalent data (the value subtracted is the one from the 1994 library). In general, the white, grey and black columns indicate differences for t, g^+ and g^- χ_i conformations respectively. Exceptions are χ_2 for His and Trp (these columns account for t, -90 and 90 and -90, 0 and 90 conformations, respectively), and χ_2 for Phe and Tyr (the difference for their only conformation is highlighted in grey). Some χ_i parameter values have not been compared, and are therefore not shown in the corresponding plot, because their χ_i dihedral angle distributions do not have well separated peaks [Asn and Asp (χ_2), Gln and Glu (χ_3) and Arg (χ_4)].

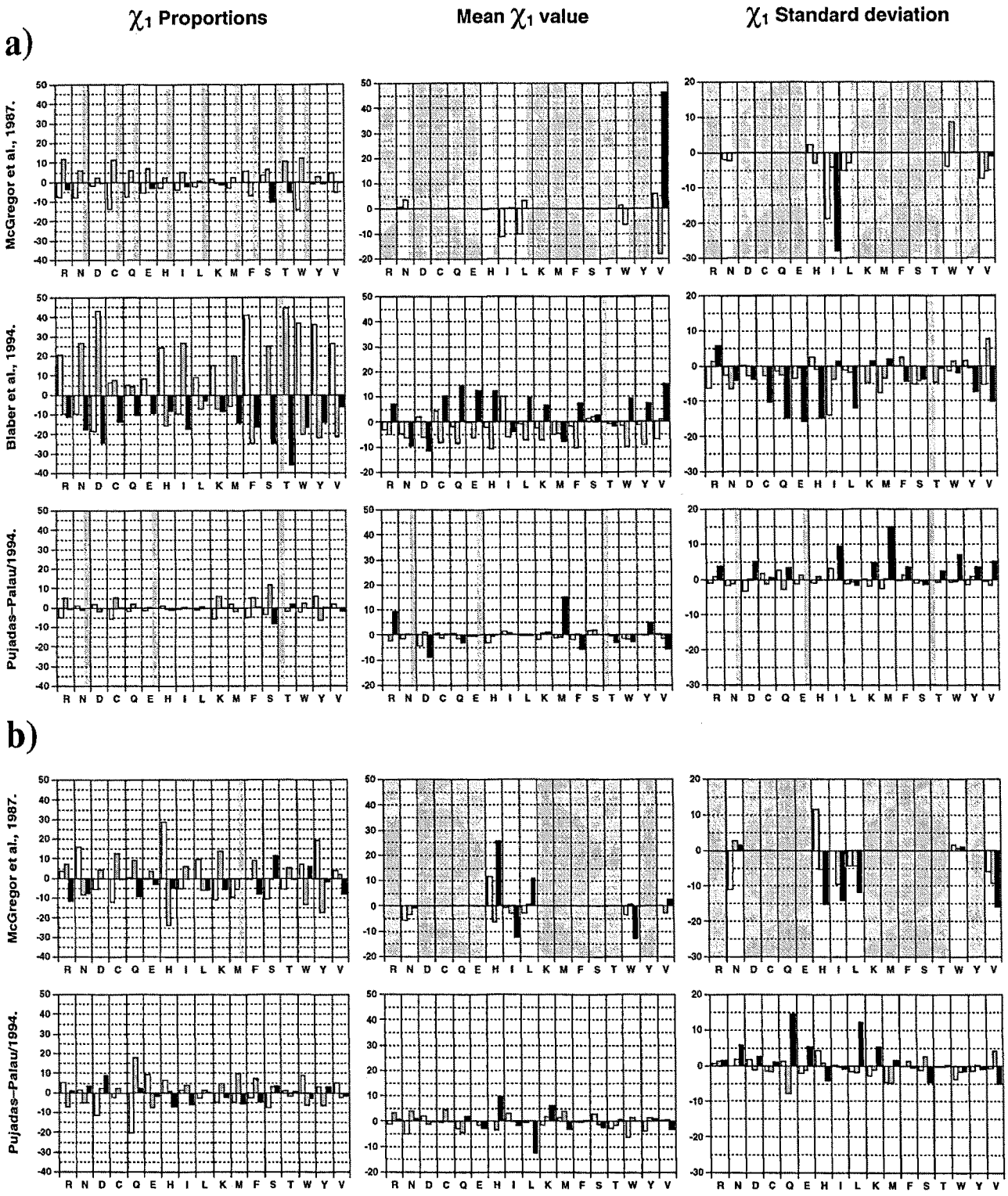


Figure 3. Comparison between the χ_1 parameters (proportion, mean value and standard deviation) found in helical N4-C4 **(a)** and β-stranded N3-C3 **(b)** rotamer libraries, and the χ_1 parameters found in Pujadas-Palau/1996 equivalent libraries. The N3-C3 segments correspond to residues that are, at least, three positions away from β-strand ends while N4-C4 segments are, at least, four positions away from helix ends. All the data used in these comparisons were taken from χ_1-based rotamer libraries. The name of the library compared with the Pujadas-Palau/1996 results is shown in the left hand margin of each row. The values shown are the differences between the data in Pujadas-Palau/1996 libraries and the equivalent data in the other libraries (the value subtracted is the one from the oldest library). The white, grey and black columns indicate differences for t, g^+ and g^- χ_1 conformations respectively. The McGregor library allows the mean χ_1 value and the standard deviation to be calculated for Asn, His, Ile, Leu, Trp and Val only. Despite this limitation, comparisons for these residues are shown as a reference. A grey background in some plots indicates that there is no data for the corresponding rotamer in at least one of the libraries used in the comparison. The data in this figure taken from the bibliography are from the α-helix centre and β-sheet centre libraries (McGregor et al., 1987) and the internal helical library (Blaber et al., 1994). These libraries coincide with the N4-C4 and N3-C3 definitions for helices and β-strands, respectively.

ISBN:978-84-694-2163-5/DL:T.885-2011

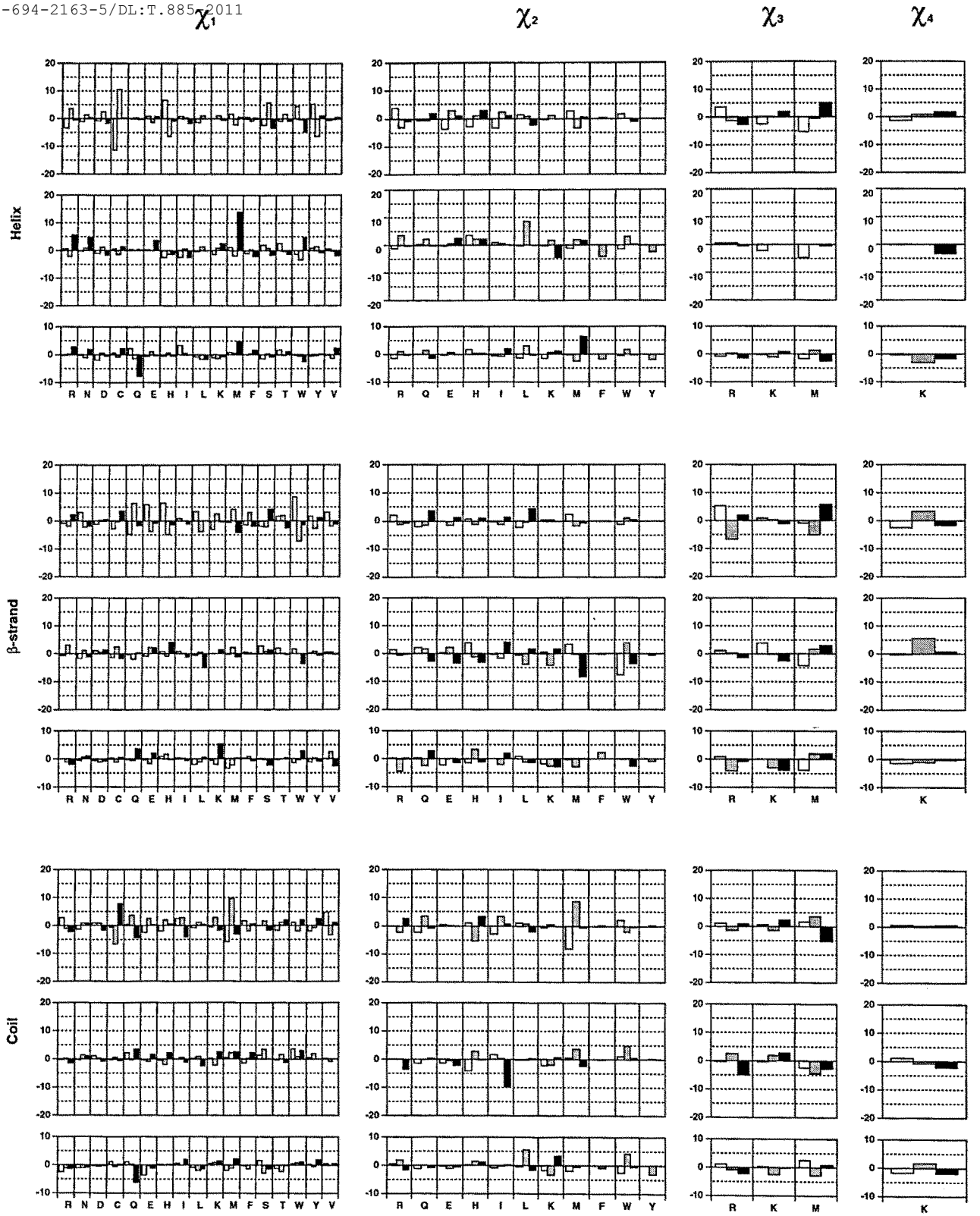


Figure 4. From top to bottom, comparisons between the χ_i parameters (proportion, mean value and standard deviation) in the χ_i-based X/Pujadas-Palau 1996 and 1994 libraries (where "X" stands for "H", "B" or "C" in helix, β-strand and coil libraries respectively). The figure shows differences between 1996 and 1994 equivalent data (the value subtracted is the one from the 1994 library). In general, the white, grey and black columns indicate differences for t, g^+ and g^- χ_i conformations respectively. Exceptions are χ_2 for His and Trp (these columns account for t, -90 and 90 and -90, 0 and 90 conformations, respectively), and χ_2 for Phe and Tyr (the difference for their only conformation is highlighted in grey). Some χ_i parameter values have not been compared, and are therefore not shown in the corresponding plot, because their χ_i dihedral angle distributions do not have well separated peaks [Asn and Asp (χ_2), Gln and Glu (χ_3) and Arg (χ_4)]. Equivalent comparisons for turns are shown in **Figure 6a**.

ISBN:978-84-694-2163-5/DL:T.885-2011

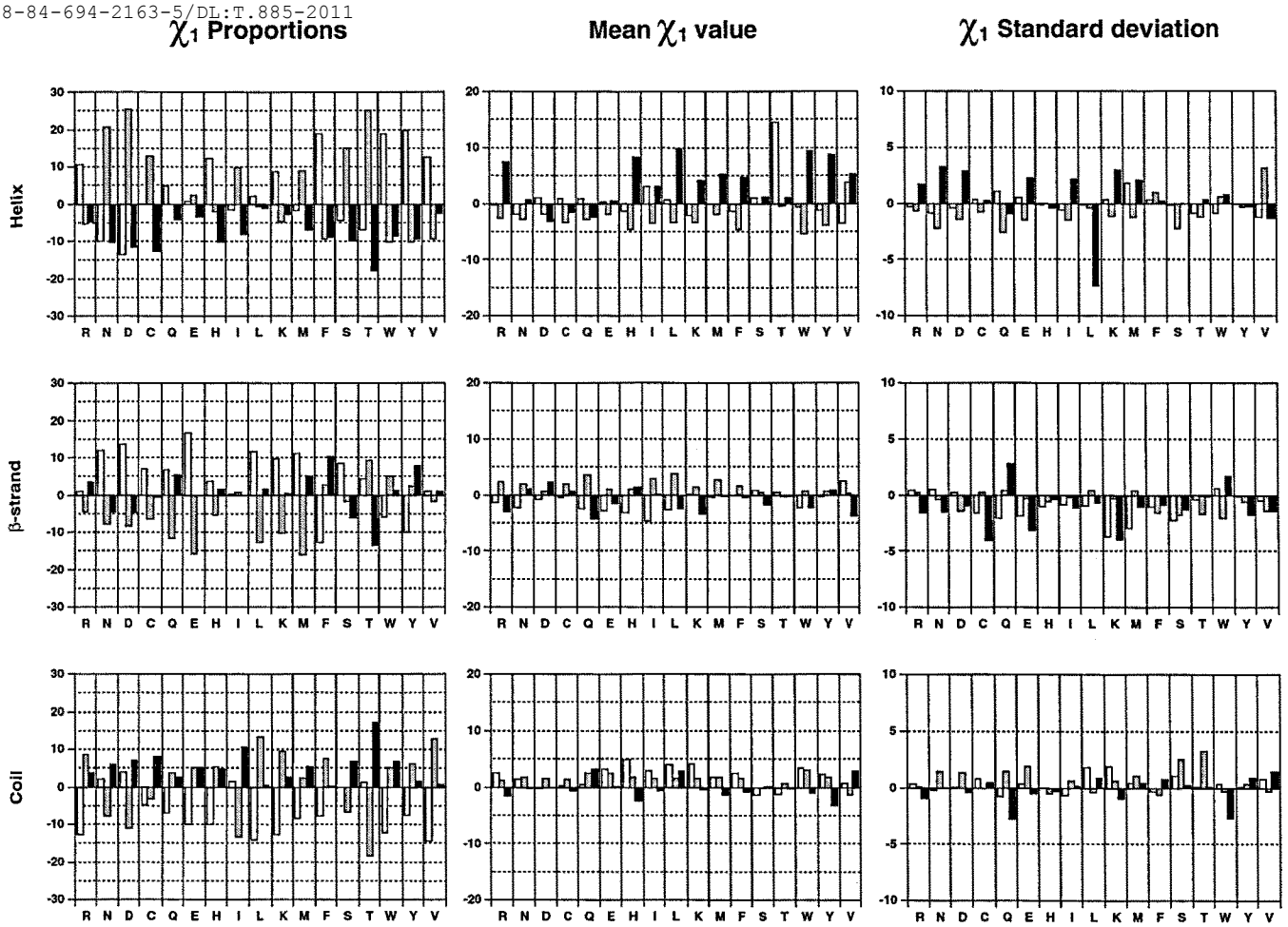
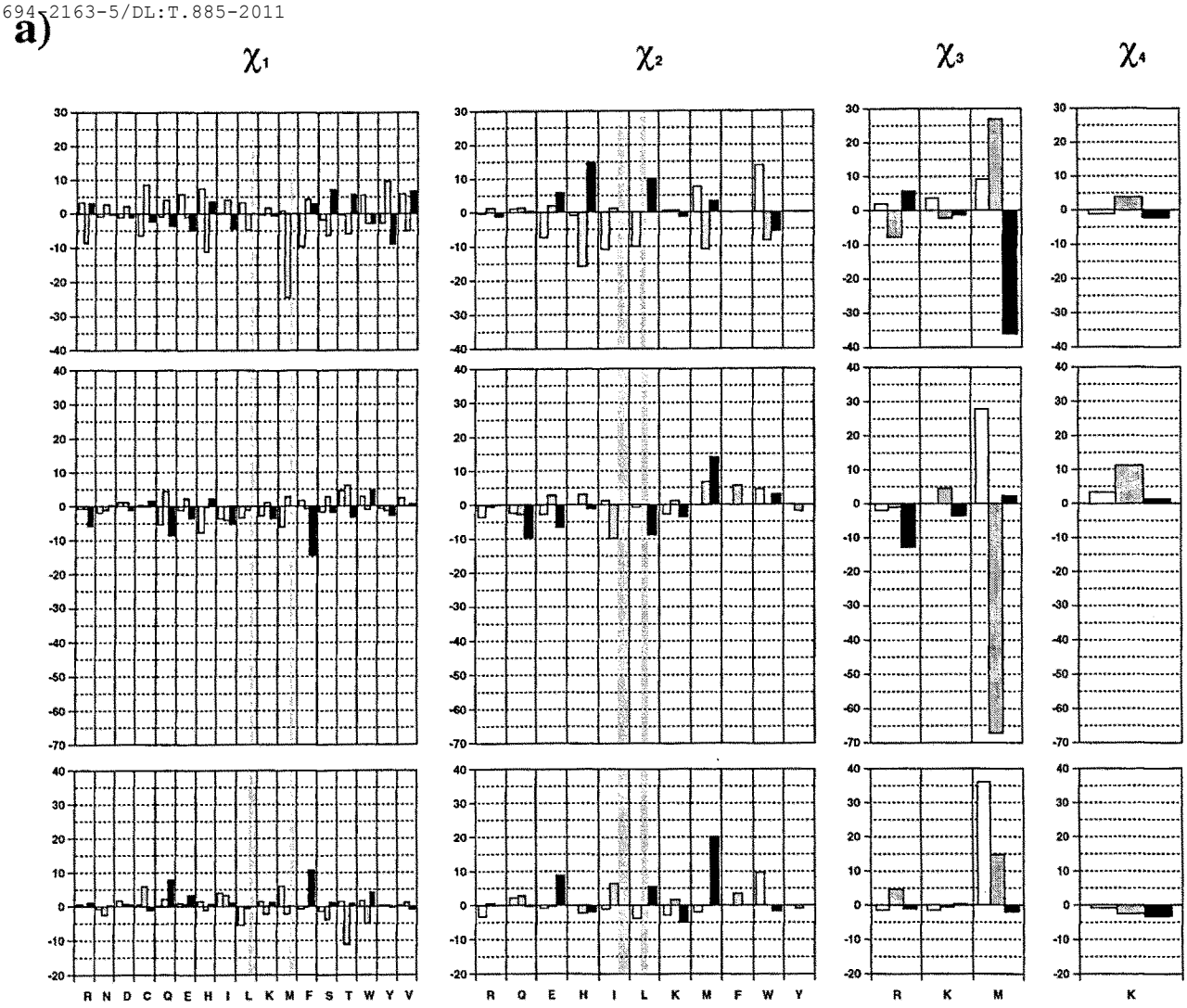


Figure 5. Comparison between the χ_1 parameters (proportion, mean value and standard deviation) in X/Pujadas-Palau/1996 (where "X" stands for "H", "B" or "C" in helix, β-strand and coil libraries respectively) and AV/Pujadas-Palau/1996. The parameters were obtained from χ_1-based libraries. The figure shows differences between H/1996, B/1996 or C/1996 and AV/1996 χ_1 parameters for the same rotamer (the value subtracted is the one from the AV library). The white, grey and black columns indicate differences for t, g^+ and g^- χ_1 conformations respectively. Equivalent comparisons for turns are shown in **Figure 6b**.

ISBN:978-84-694-2163-5/DL:T.885-2011



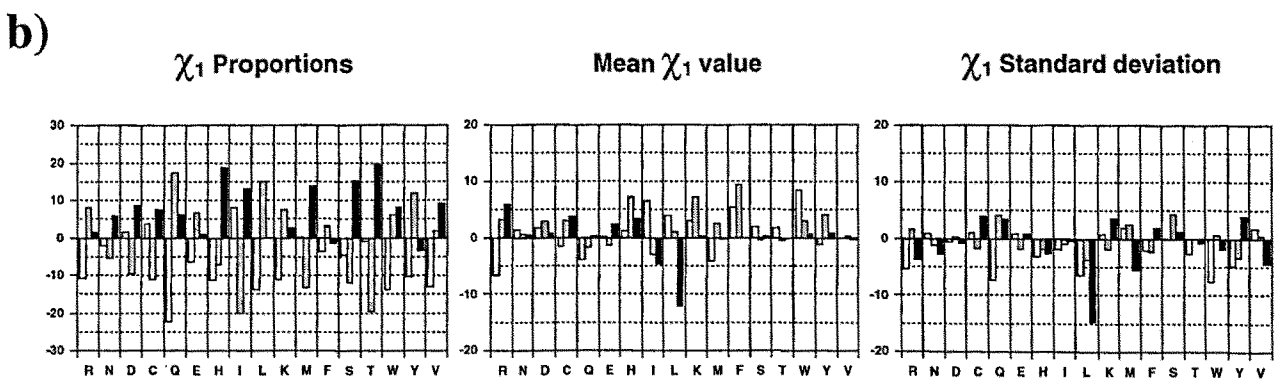


Figure 6. Results for turns. **a)** Equivalent to **Figure 4**. From top to bottom, comparisons between the χ_1 parameters (proportion, mean value and standard deviation) in the χ_1-based T/Pujadas-Palau 1996 and 1994 libraries. The figure shows differences between 1996 and 1994 equivalent data (the value subtracted is the one from the 1994 library). Some χ_1 parameter values have not been compared, and are therefore not shown in the corresponding plot, because their χ_1 dihedral angle distributions do not have well separated peaks [Asn and Asp (χ_2), Gln and Glu (χ_3) and Arg (χ_4)]. A grey background in some plots indicates that there is no equivalent data in the 1994 library. **b)** Equivalent to **Figure 5**. Comparison between the χ_1 parameters (proportion, mean value and standard deviation) in T/Pujadas-Palau/1996 and AV/Pujadas-Palau/1996. The parameters were obtained from χ_1-based libraries. The figure shows differences between T/1996 and AV/1996 χ_1 parameters for the same rotamer (the value subtracted is the one from the AV library). In general, the white, grey and black columns indicate differences for t, g^+ and g^- χ_1 conformations respectively. Exceptions are χ_2 for His and Trp (these columns account for t, -90 and 90 and -90, 0 and 90 conformations, respectively), and χ_2 for Phe and Tyr (the difference for their only conformation is highlighted in grey).

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8: RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

Taules de l'apèndix.

Table 1. Compilation of rotamer libraries used in programs for modelling or predicting protein side-chain conformations.

Ponder & Richards (1987)	**McGregor et al. (1987)**	**Tuffery et al. (1991)**	**Desmet et al. (1992) ***
Tuffery et al. (1997)	Desjarlais & Handel (1995)	Tuffery et al. (1997)	De Maeyer et al. (1997)
Dahiyat & Mayo (1996)		Desjarlais & Handel (1995)	Lasters et al. (1995)
Nakamura et al. (1996)		Keller et al. (1995)	Goldstein (1994)
Shenkin et al. (1996)		Koehl & Delarue (1995)	Leach (1994)
Castonguay et al. (1995)		Selvaggini et al. (1995)	Tanimura et al. (1994)
Harbury et al. (1995)		Cregut et al. (1994)	
Desmet et al. (1994)		Goldstein (1994)	
Hsieh et al. (1994)		Koehl & Delarue (1994)	
Kono & Doi (1994)		Tuffery et al. (1994)	
Tanaka et al. (1994)		van Gelder et al. (1994)	
Tanimura et al. (1994)			

Dunbrack & Karplus (1993)	**Tuffery/1995**	**Dunbrack/1996**
Iyer & Vishveshwara (1996)	Tuffery et al. (1997)	Bower et al. (1997)

The table shows which rotamer libraries were used by the different authors for modelling or predicting side-chain conformation. Only references published since 1994 have been compiled. *The Desmet et al. (1992) rotamer library is an extended version of the Ponder & Richards (1987) library.

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8:
RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

Table 2. Comparison of χ_1 proportions from all known averaged rotamer libraries which were developed before 1997.

	Pujadas-Palau/1996	Dunbrack/1996	Pujadas-Palau/1994	Blaber et al. (1994)	Dunbrack (1993)	Dunbrack & Karplus (1993)	Tuffery et al. (1991)	Ponder & Richards (1987)	McGregor et al. (1987)
Pujadas-Palau/1996		Trp g-	Cys t	Gln t,g+		Cys g+	Arg g+; **Ile g+**,g- Met g+ ***Phe*** **t,*g+*** **,*g-*** ***Trp*** **g+** **,g-**	Arg t; Asn t,g+; Cys g+ Glu g+; **Ile t,g+** ***Met*** **t,*g+*** Phe t,g-; **Trp t,g-** Val g+	Arg g-; Asn t Gln t,g+; Glu t,g+ Ile g+; Leu g+ **Met t,g+** Thr g-; Val g-
Dunbrack/1996	1.0 1.2 1.3 1.2		Trp g+,g-	Gln t,g+	Trp g+,g-	Trp g+,g-	Arg g+; *His t,g+* Ile g+,g- ***Phe*** **t,*g+*** **,*g-*** **Trp t,g+**	Asn t,g+; Glu g+ His g+; Ile t,g+ **Met t,g+** Phe t,g- Trp t; Val g+	Gln t,g+; Glu g+ Ile g+ **Met t,g+** Thr g-; Trp g-
Pujadas-Palau/1994	1.5 1.5 0.9 1.3	1.1 1.8 1.4 1.4		Gln t		Cys t,g+	Arg g+; Asp g+; Cys t,g+ Gln g+,g-; His t **Ile g+**,g-; Met t,g+ ***Phe*** **t,*g+*** **,*g-*** ; Thr g+ ***Trp*** **t,*g+*** ,g-; Tyr g+	Arg t; Asn t,g+ Cys t,g+; Glu g+,g- **Ile t,g+** ***Met*** **t,*g+*** ; Phe t,g- **Trp t,g-**; Val g+	Gln t; Glu t,g+ Ile t,g+; Leu g+ Met t,g+; Thr g- Trp t,g+; Val g-
Blaber et al. (1994)	1.7 1.8 1.2 1.5	1.7 1.7 0.8 1.4	2.0 2.0 0.9 1.6			Glu g+	Arg g+; Gln **t,g+** Glu t,g+ Ile g+,g-; Met t,g+ ***Phe*** **g+** **,g-** **Trp** t,g+,g-	Asn t,g+; Asp t Glu t,g+; Ile g+ ***Met*** **t,*g+*** **Phe t,g-** Trp t,g-	Asn t; Glu t,g+ Met t,g+; Thr g-
Dunbrack (1993)	1.4 1.5 1.5 1.4	1.2 1.4 1.1 1.2	1.5 1.5 1.1 1.4	1.4 1.3 0.9 1.2			Arg t,g+; Asp g+ Gln g+,g-; Ile g+,g- Met t,g+ ***Phe*** **t,*g+*** **,*g-*** ***Trp*** **t,*g+*** ,g-	Asn t,g+; Glu g+ Ile t,g+; ***Met*** **t,*g+*** Phe t,g- **Trp t,g-** Val g+	Glu g+; Leu g+ Met t,g+; Thr g- Tyr t
Dunbrack & Karplus (1993)	1.4 2.2 1.6 1.7	1.4 2.0 1.2 1.5	1.7 1.7 1.2 1.5	1.4 1.5 1.0 1.3	0.8 0.9 0.3 0.7		Arg t,g+; Asn g+ Gln g+,g-; His g+ Ile g+,g-; Met t,g+ ***Phe*** **g+** **,g-** ***Trp*** **t,*g+*** ,g-	Asn t,g+; His g+ Ile t,g+ ***Met*** **t,*g+*** **Phe t,g-** **Trp t,g-**	Asn t; Met t,g+ Thr g-
Tuffery et al. (1991)	2.9 7.2 5.2 5.1	3.3 7.0 4.3 4.8	4.0 8.2 5.0 5.7	3.4 7.3 4.4 5.1	3.5 7.4 4.3 5.1	3.4 7.4 4.5 5.1		Arg t,g+; **Asn t,g+** Asp g+; Gln g+; Leu t,g+ Lys g+; Met g+ ***Phe*** **t,*g+*** **,*g-*** ; Thr g- ***Trp*** **t** **,g+**; Val g+	Arg t; Asn t,g+ **Gln g+**,g- His t; **Met t,g+** ***Phe*** **t,*g+*** **,*g-*** **Trp g+**,g-; Tyr t,g+
Ponder & Richards (1987)	4.1 4.3 4.0 4.1	3.7 4.3 3.3 3.8	4.2 4.8 4.0 4.3	4.1 4.3 3.6 4.0	3.5 3.8 3.8 3.7	3.6 4.2 3.9 3.9	4.9 8.0 4.8 5.9		Asn g+,g-; Leu g+ Lys g+; ***Met*** **t** **,*g+*** **,g-** Phe t,g- Ser g-; Thr t,g- **Trp t,g-**; Val g+,g-
McGregor et al. (1987)	3.5 3.9 3.0 3.5	3.3 3.5 2.7 3.2	3.5 3.5 2.6 3.2	3.1 3.1 2.5 2.9	2.8 3.2 2.1 2.7	3.1 2.9 2.2 2.7	4.5 6.8 4.7 5.3	4.5 5.3 4.3 4.7	

The table shows the results of all possible cross comparisons of χ_1 proportions from all known averaged rotamer libraries which were developed before 1997. The libraries were obtained from bibliographic references, Internet and our own results. In all cases, the data used for comparisons come from χ_1-based libraries. For each pair of libraries, the data above the diagonal, if any, shows all conformations with differences in their χ_1 proportions beyond the ±5% interval. If differences are beyond ±10% the conformations are in bold and if they are beyond ±15% in bold italics. The amino acids are also in bold and bold italics to identify the distant outliers more easily. Values below the diagonal show the "Relative Quality Factor (RQF)" which is defined as the average of the absolute values for differences in the χ_1 proportions. From top to bottom, RQF values for the t, g^+, g^- and global (t, g^+, g^- pooled together) differences are shown for each pair of libraries.

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8:
RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Aymerich
ISBN:978-84-693-1135/DL:T.885-2011

Table 3. Comparison of mean χ_1 values and their standard deviations from all known averaged rotamer libraries which were developed before 1997.

a)

	Pujadas-Palau/1996	Dunbrack/1996	Pujadas-Palau/1994	Blaber et al. (1994)	Dunbrack (1993)	Tuffery et al. (1991)	Ponder & Richards (1987)
Pujadas-Palau/1996				Ile g- Thr t Val g+		Arg g-; Ile t Leu g-; ***Met g-*** **Thr t**; Val g-	**Arg g-**; Asn t Gln g-; Glu g- ***Leu g-***; Lys g- Met t
Dunbrack/1996	0.7 1.4 1.4 1.2		Leu g-	Ile g-		**Arg g-**; Gln g- His g-; Ile t Lys g-; ***Met g-*** **Thr t**; Val g-	**Arg g-**; Asn t Gln g-; Glu g- His g-; ***Leu g-*** Lys g-; Met t Ser g+; Val g-
Pujadas-Palau/1994	0.6 0.6 1.1 0.8	0.8 1.8 1.9 1.5		His g-; Thr t Val g+, g-	Leu g-	Arg g-; Glu g- His g-; Ile t Leu g-; Lys g- ***Met g-***; **Thr t** Val g-	**Arg g-**; Asn t Gln g-; Glu g- His g-; ***Leu g+, g-*** Lys g-; Met t
Blaber et al. (1994)	1.6 1.1 1.8 1.5	1.4 1.5 2.1 1.7	1.6 1.2 1.9 1.6		Leu g-	**Arg g-**; Gln g- Glu g-; Ile t Leu g-; Lys g- ***Met g-***; ***Thr t*** **Val g+, g-**	**Arg g-**; Asn t Gln g-; Glu g- ***Leu g-***; Lys g- Met t; Thr t Val g+, g-
Dunbrack (1993)	1.5 0.8 2.0 1.4	1.3 1.3 1.5 1.4	1.4 0.9 2.1 1.5	1.0 0.9 1.6 1.2		**Arg g-**; Gln g- Glu g-; Ile t Lys g-; ***Met g-*** **Thr t**; Val g-	**Arg g-**; Asn t Gln g-; Glu g- ***Leu*** g+, ***g-***; **Lys g-** **Met t**; Thr t Val g-
Tuffery et al. (1991)	2.5 1.2 4.3 2.7	2.5 1.4 5.0 2.9	2.5 1.3 4.9 2.9	2.9 1.7 4.5 3.0	2.6 1.2 4.4 2.7		Asn t; ***Leu*** g+, ***g-*** Met t; Ser g+ Thr t
Ponder & Richards (1987)	2.0 1.9 4.7 2.8	2.1 1.7 5.5 3.1	2.3 1.9 5.3 3.1	3.1 2.0 5.4 3.5	2.8 2.1 5.1 3.3	3.0 1.9 2.6 2.5	

b)

	Pujadas-Palau/1996	Dunbrack/1996	Pujadas-Palau/1994	Blaber et al. (1994)	Dunbrack (1993)	Ponder & Richards (1987)
Pujadas-Palau/1996				His g-; **Ile t** Lys g-; Met t Ser t; **Thr t** **Val g-**	Gln g-; **Ile t**, g- Lys g-; Met t Ser t; Thr t **Val g+, g-**	Arg g-; Asn g- Glu g-; His g- Leu g+, g-; Ser t
Dunbrack/1996	1.4 1.9 1.1 1.4			Arg t; Glu t, g- **Ile t**; Lys t, g+, g- Met t; Ser t Thr t; **Val g-**	Arg t; Gln t, g- Glu t, g+, g-; **Ile t**, g- Leu t; Lys t, g+, g- Met t; Ser t Thr t; **Val g+, g-**	Arg g-; Asn g- Glu g+, g-; His g- Leu g-; Lys t, g+ Ser t; Thr t
Pujadas-Palau/1994	0.6 0.6 0.6 0.6	1.5 2.2 1.2 1.6		His g-; **Ile t** Lys g-; Ser t **Thr t; Val g-**	Gln t; **Ile t**, g- Lys g-; Ser t Thr t; **Val g+, g-**	Arg g-; Asn g- Gln g-; Glu g- His g-; Leu g+, g- Ser t
Blaber et al. (1994)	3.4 0.9 2.9 2.4	4.3 2.1 3.3 3.2	3.2 0.8 2.9 2.3			Asn g-; **Glu g-** Ile t, g-; Leu t, g+ Lys g-; Met t ***Thr t***; Trp t ***Val g-***
Dunbrack (1993)	3.7 1.5 3.8 3.0	4.7 3.3 4.2 4.1	3.4 1.3 3.5 2.7	0.9 1.3 1.3 1.2		Asn g-; Asp t; Gln g- **Glu g-**; **Ile t**, g- Leu t, g+; Lys g- Met t; **Thr t** Trp t, ***Val*** g+, ***g-***
Ponder & Richards (1987)	2.3 1.7 3.6 2.5	2.9 2.0 3.4 2.7	2.7 2.0 3.6 2.7	3.8 2.0 4.7 3.5	4.1 2.9 6.1 4.3	

he table shows the results of all possible cross comparisons of mean χ_1 values **(a)** and their standard deviations **(b)** from all known averaged rotamer libraries which were developed efore 1997. The libraries were obtained from bibliographic references, Internet and our own results. In all cases, the data used for comparisons come from χ_1-based libraries. For ich pair of libraries, the data above the diagonal, if any, shows all conformations with differences in their χ_1 parameters beyond the ±5° interval. If differences are beyond ±10° the nformations are in bold and if they are beyond ±15° in bold italics. The amino acids are also in bold and bold italics to identify distant outliers more easily. Values below the agonal show the "Relative Quality Factor (RQF)" which is defined as the average of the absolute values of differences in the mean χ_1 values **(a)** and in their standard deviations **(b)**. om top to bottom, RQF values for the t, g^+, g^- and global (t, g^+, g^- pooled together) differences are shown for each pair of libraries.

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8: RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

Table 4. Codes of the subunits used for building the 1996 and 1994 rotamer library sets.

(a)

1ABA	1EDE	1LTSA	1SBP	2CTC	2TGI	4BLMA
1ARB	**1EZM**	1LTSC	1SNC	2DNJA	3B5C	4DFRA
1BBHA	1FBAA	1LTSD	**1TRB**	2END	3CHY	4ENL
1BBPA	**1GKY**	1MDC	1TTBA	2ER7E	3CLA	4FGF
1BGH	1GOF	1NAR	2ABK	2FAL	3COX	4FXN
1CCR	1GP1A	1ONC	**2ALP**	2HPDA	3DFR	4GCR
1CEWI	1GPB	1PAZ	2AYH	2MNR	3GRS	5P21
1CMBA	1HBQ	1PDA	2AZAA	2PIA	3RUBL	5TIMA
1CPCA	1HUW	1PII	2BBKH	2POR	3SDHA	8ABP
1CSEI	1ISUA	1POC	2CCYA	2RSPA	3SGBI	8ACN
1CTF	1LENA	1POH	2CDV	2SCPA	**3TGL**	9RNT
1DSBA	1LIS	**1REC**	2CPL			

(b)

193L (**2IHL**)	**1CPCB** (**1CPCL**)	1FSC (1FAS)	1MLS (1MGN)	1RTP1 (1RRO)	1TYS (2TSCA)	**2SIL** (**2SIM**)
1AAC (1AAJ)	1CSEE (**1S01**)	1GSA (2GLT)	1MLS (2SPO)	**1SLTB** (**1SLTA**)	2ACQ (**1ADS**)	7PTI (1AAPA)
1BRIC (1BSAA)	1CSH (2CTS)	1JCV (1COBA)	1PTF (2HPR)	**1SRIB** (**1PTSA**)	2CBA (1CAJ)	8TLNE (6TMNE)
1BYB (1BTC)	**1FIPA** (1FIAB)	1L17 (129L)	1RBU (2RN2)	1TOP (1OSA)	2CWGA (9WGAA)	9PAP (**1PPN**)

(c)

Non-equivalent subunits used in the 1996 sample

153L	1DIN	1GLQA	1LCT	1PHG	**1TIB**	2HMZA
1ABOA	1DOI	1GPR	1LENB	1PNE	1TRKA	2HTS
1AMP	1DPE	1GSEA	1LKI	1PNKA	1TSP	2KAUA
1ARV	1DSUA	1HAN	1LMB3	1PNKB	1UBI	2KAUB
1ATLA	1DTS	1HGXA	1MJC	1PRN	1UBSA	2KAUC
1BAM	1DUPA	1HMT	1MKAA	1PTQ	1UBSB	2NACA
1BDMB	1DYR	1HOE	1MLA	1PTX	1UKZ	2OLBA
1BDO	1ECPA	**1HSLA**	1MMD	1QUK	1VID	2ORA
1BP2	1EDG	1HVD	1MML	1R69	1VSD	2PGD
1BRSD	1EDT	1HXN	1MOLA	1RCI	1WAPA	2PHY
1CELA	1ENP	1I1B	1MSC	1REGX	1WDCA	2PII
1CEO	1ESL	**1IAE**	1MUCA	1RIS	1WHI	2PRD
1CGO	1FIVA	**1IGD**	**1NBAA**	**1RVAA**	1XNB	2PRK
1CHD	1FKJ	1IGS	1NFP	1SACA	1XYZA	3BCL
1CHMA	1FLP	1ILK	1NHKL	1SAT	256BA	**3PGA1**
1CLXA	1FMB	1ISCA	1NHP	1SLUA	2AK3A	3PTE
1CNSA	1FNC	1JAPA	1NIF	1SMD	2CY3	3SICI
1CNV	**1FRD**	1KPBA	1OACA	1SRA	2EBN	451C
1CTJ	1FUA	1KPTA	1PBE	1SVB	2FX2	4ICB
1CYDA	1GADO	1LATA	1PBN	1TCA	2GDM	4MT2
1DAAA	1GARA	1LAUE	1PDO	1TGXA	2GSTA	5RUBA
1DDT	1GDOA	1LCPA	**1PGS**	**1THV**	2HBG	**8FABB**

Non-equivalent subunits used in the 1994 sample

1ALKA	1DFNA	1HIVA	**1PPA**	**1SHG**	**2MSBA**	4XIS
1APME	1GD1O	1OFV	1PPFE	1TGSI	**3IL8**	8DFR
1BABB	1GMPA	1OPBA	1RND	1TROA	3RUBS	**8RXNA**
1CGT	1GOX	1POA	**1SHFA**			

The subunit name is formed by the corresponding BPDB four-letter code followed by the subunit label used in the BPDB file (if any). The codes in bold refer to chains which have no information about secondary-structure limits in the original BPDB files (in these cases, the information was taken from the corresponding DSSP entry). **(a)** Subunits common to 1996 and 1994 samples. **(b)** Subunits in the 1996 sample that are equivalent to others used to develop the 1994 libraries (in brackets). In this case, the 1996 structure generally belongs to the same protein (but not necessarily to the same source) as the 1994 chain but usually has better structural quality. **(c)** Non-equivalent subunits for each sample.

Table 5. Distribution of residues in the samples used for building the 1996 and 1994 rotamer library sets.

	Helix				β-strand				Turn				Coil				Total			
	96/94				96/94				96/94				96/94				96/94			
Ala	1889	[12.8]	903	[13.4]	705	[6.9]	305	[6.6]	118	[6.6]	70	[5.9]	1080	[6.4]	455	[6.6]	3792	[8.7]	1733	[8.9]
Arg	851	[5.8]	366	[5.4]	408	[4.0]	182	[3.9]	48	[2.7]	41	[3.4]	686	[4.1]	290	[4.2]	1993	[4.6]	879	[4.5]
Asn	561	[3.8]	253	[3.8]	303	[3.0]	147	[3.2]	132	[7.3]	99	[8.3]	1111	[6.6]	438	[6.3]	2107	[4.8]	937	[4.8]
Asp	783	[5.3]	343	[5.1]	363	[3.5]	186	[4.0]	181	[10.1]	124	[10.4]	1348	[8.0]	528	[7.7]	2675	[6.1]	1181	[6.1]
Cys	173	[1.2]	91	[1.4]	157	[1.5]	78	[1.7]	40	[2.2]	22	[1.8]	253	[1.5]	109	[1.6]	623	[1.4]	300	[1.5]
Gln	704	[4.8]	284	[4.2]	241	[2.3]	121	[2.6]	53	[2.9]	36	[3.0]	550	[3.3]	197	[2.9]	1548	[3.5]	638	[3.3]
Glu	1160	[7.9]	532	[7.9]	456	[4.4]	193	[4.2]	94	[5.2]	71	[6.0]	788	[4.7]	328	[4.8]	2498	[5.7]	1124	[5.8]
Gly	586	[4.0]	300	[4.5]	589	[5.7]	264	[5.7]	299	[16.6]	204	[17.1]	2137	[12.6]	838	[12.1]	3611	[8.3]	1606	[8.3]
His	286	[1.9]	143	[2.1]	243	[2.4]	97	[2.1]	36	[2.0]	27	[2.3]	402	[2.4]	191	[2.8]	967	[2.2]	458	[2.4]
Ile	825	[5.6]	358	[5.3]	980	[9.5]	429	[9.2]	28	[1.6]	17	[1.4]	615	[3.6]	239	[3.5]	2448	[5.6]	1043	[5.4]
Leu	1564	[10.6]	693	[10.3]	1002	[9.8]	436	[9.4]	75	[4.2]	52	[4.4]	978	[5.8]	401	[5.8]	3619	[8.3]	1582	[8.1]
Lys	1008	[6.8]	472	[7.0]	484	[4.7]	214	[4.6]	113	[6.3]	85	[7.1]	943	[5.6]	384	[5.6]	2548	[5.8]	1155	[5.9]
Met	388	[2.6]	175	[2.6]	235	[2.3]	105	[2.3]	17	[0.9]	7	[0.6]	226	[1.3]	83	[1.2]	866	[2.0]	370	[1.9]
Phe	620	[4.2]	277	[4.1]	538	[5.2]	265	[5.7]	48	[2.7]	27	[2.3]	580	[3.4]	255	[3.7]	1786	[4.1]	824	[4.2]
Pro	303	[2.1]	138	[2.1]	221	[2.2]	107	[2.3]	122	[6.8]	75	[6.3]	1269	[7.5]	498	[7.2]	1915	[4.4]	818	[4.2]
Ser	735	[5.0]	343	[5.1]	524	[5.1]	267	[5.8]	157	[8.7]	90	[7.6]	1218	[7.2]	489	[7.1]	2634	[6.0]	1189	[6.1]
Thr	654	[4.4]	305	[4.5]	728	[7.1]	344	[7.4]	108	[6.0]	60	[5.0]	1129	[6.7]	512	[7.4]	2619	[6.0]	1221	[6.3]
Trp	222	[1.5]	99	[1.5]	196	[1.9]	84	[1.8]	18	[1.0]	12	[1.0]	227	[1.3]	101	[1.5]	663	[1.5]	296	[1.5]
Tyr	506	[3.4]	211	[3.1]	538	[5.2]	242	[5.2]	53	[2.9]	33	[2.8]	543	[3.2]	233	[3.4]	1640	[3.8]	719	[3.7]
Val	905	[6.1]	433	[6.4]	1356	[13.2]	576	[12.4]	60	[3.3]	40	[3.4]	825	[4.9]	332	[4.8]	3146	[7.2]	1381	[7.1]
Total	14723		6719		10267		4642		1800		1192		16908		6901		43698		19454	

The table shows how the residues used to build both rotamer library sets are distributed among helices, β-strands, turns, coils and in the global sample. For each structure type, the first and the second pair of columns show data for the 1996 and 1994 samples, respectively. For each pair of columns, the first shows the number of residues for every amino acid whereas the second shows their proportion within the corresponding structure (in square brackets). The "Total" row indicates the total number of residues in each structural type. Data shown in the "Total" columns were obtained after pooling residues for the same amino acid, independently of their structural origin.

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8:
RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

Biblioteques de rotàmers.

ISBN:978-84-694-2163-5/DL:T.885-2011

Appendix Table 1. Distribution of residues among rotamers for dependent and nondependent secondary-structure "full rotamer libraries".

Residue	Rotamer label *	Distribution *														
		Averaged		Helix			β-strand			Turn			Coil		
Arg	t t t	203	10.2	89	10.5	43.8	52	12.8	25.6	6	12.5	3.0	56	8.2	27.6
	t t g+	171	8.6	87	10.2	50.9	48	11.8	28.1	-	-	-	36	5.3	21.1
	t t g-	172	8.6	112	13.2	65.1	29	7.1	16.9	2	4.2	1.2	29	4.2	16.9
	t g+ t	8	0.4	5	0.6	62.5	-	-	-	-	-	-	3	0.4	37.5
	t g+ g+	8	0.4	6	0.7	75.0	1	0.3	12.5	-	-	-	1	0.2	12.5
	t g+ g-	1	0.1	1	0.1	100.0	-	-	-	-	-	-	-	-	-
	t g- t	65	3.3	43	5.1	66.2	6	1.5	9.2	2	4.2	3.1	14	2.0	21.5
	t g- g+	7	0.4	4	0.5	57.1	1	0.3	14.3	-	-	-	2	0.3	28.6
	t g- g-	30	1.5	23	2.7	76.7	5	1.2	16.7	1	2.1	3.3	1	0.2	3.3
	g+ t t	405	20.3	192	22.6	47.4	74	18.1	18.3	7	14.6	1.7	132	19.2	32.6
	g+ t g+	272	13.7	115	13.5	42.3	39	9.6	14.3	7	14.6	2.6	111	16.2	40.8
	g+ t g-	199	10.0	71	8.3	35.7	32	7.8	16.1	7	14.6	3.5	89	13.0	44.7
	g+ g+ t	113	5.7	21	2.5	18.6	41	10.1	36.3	4	8.3	3.5	47	6.9	41.6
	g+ g+ g+	68	3.4	19	2.2	27.9	14	3.4	20.6	5	10.4	7.4	30	4.4	44.1
	g+ g+ g-	18	0.9	5	0.6	27.8	2	0.5	11.1	1	2.1	5.6	10	1.5	55.6
	g+ g- t	21	1.1	4	0.5	19.1	3	0.7	14.3	-	-	-	14	2.0	66.7
	g+ g- g+	5	0.3	3	0.4	60.0	1	0.3	20.0	-	-	-	1	0.2	20.0
	g+ g- g-	11	0.6	1	0.1	9.1	5	1.2	45.5	-	-	-	5	0.7	45.5
	g- t t	88	4.4	16	1.9	18.2	27	6.6	30.7	3	6.3	3.4	42	6.1	47.7
	g- t g+	39	2.0	11	1.3	28.2	7	1.7	18.0	2	4.2	5.1	19	2.8	48.7
	g- t g-	37	1.9	6	0.7	16.2	15	3.7	40.5	-	-	-	16	2.3	43.2
	g- g+ t	2	0.1	1	0.1	50.0	-	-	-	-	-	-	1	0.2	50.0
	g- g+ g+	2	0.1	-	-	-	-	-	-	-	-	-	2	0.3	100.0
	g- g+ g-	1	0.1	-	-	-	-	-	-	-	-	-	1	0.2	100.0
	g- g- t	6	0.3	1	0.1	16.7	1	0.3	16.7	-	-	-	4	0.6	66.7
	Other	41	2.1	15	1.8	36.6	5	1.2	12.2	1	2.1	2.4	20	2.9	48.8
Asn	t	602	28.6	104	18.5	17.3	123	40.6	20.4	35	26.5	5.8	340	30.6	56.5
	g+	1149	54.5	422	75.2	36.7	142	46.9	12.4	65	49.2	5.7	520	46.8	45.3
	g-	341	16.2	33	5.9	9.7	35	11.6	10.3	29	22.0	8.5	244	22.0	71.6
	Other	15	0.7	2	0.4	13.3	3	1.0	20.0	3	2.3	20.0	7	0.6	46.7
Asp	t	816	30.5	133	17.0	16.3	160	44.1	19.6	58	32.0	7.1	465	34.5	57.0
	g+	1352	50.5	594	75.9	43.9	153	42.2	11.3	74	40.9	5.5	531	39.4	39.3
	g-	492	18.4	53	6.8	10.8	49	13.5	10.0	49	27.1	10.0	341	25.3	69.3
	Other	15	0.6	3	0.4	20.0	1	0.3	6.7	-	-	-	11	0.8	73.3
Cys	t	163	26.2	45	26.0	27.6	52	33.1	31.9	12	30.0	7.4	54	21.3	33.1
	g+	349	56.0	119	68.8	34.1	78	49.7	22.4	18	45.0	5.2	134	53.0	38.4
	g-	108	17.3	8	4.6	7.4	26	16.6	24.1	10	25.0	9.3	64	25.3	59.3
	Other	3	0.5	1	0.6	33.3	1	0.6	33.3	-	-	-	1	0.4	33.3
Gln	t t	279	18.0	129	18.3	46.2	61	25.3	21.9	2	3.8	0.7	87	15.8	31.2
	t g+	34	2.2	20	2.8	58.8	4	1.7	11.8	-	-	-	10	1.8	29.4
	t g-	147	9.5	94	13.4	64.0	23	9.5	15.7	2	3.8	1.4	28	5.1	19.1
	g+ t	630	40.7	313	44.5	49.7	86	35.7	13.7	23	43.4	3.7	208	37.8	33.0
	g+ g+	263	17.0	99	14.1	37.6	28	11.6	10.7	11	20.8	4.2	125	22.7	47.5
	g+ g-	65	4.2	22	3.1	33.9	7	2.9	10.8	8	15.1	12.3	28	5.1	43.1
	g- t	77	5.0	12	1.7	15.6	22	9.1	28.6	5	9.4	6.5	38	6.9	49.4
	g- g+	21	1.4	5	0.7	23.8	5	2.1	23.8	-	-	-	11	2.0	52.4
	g- g-	13	0.8	3	0.4	23.1	3	1.2	23.1	2	3.8	15.4	5	0.9	38.5
	Other	19	1.2	7	1.0	36.8	2	0.8	10.5	-	-	-	10	1.8	52.6
Glu	t t	564	22.6	258	22.2	45.7	184	40.4	32.6	16	17.0	2.8	106	13.5	18.8
	t g+	38	1.5	18	1.6	47.4	5	1.1	13.2	1	1.1	2.6	14	1.8	36.8
	t g-	194	7.8	103	8.9	53.1	32	7.0	16.5	6	6.4	3.1	53	6.7	27.3
	g+ t	861	34.5	394	34.0	45.8	149	32.7	17.3	32	34.0	3.7	286	36.3	33.2
	g+ g+	374	15.0	194	16.7	51.9	34	7.5	9.1	22	23.4	5.9	124	15.7	33.2
	g+ g-	186	7.5	98	8.5	52.7	5	1.1	2.7	6	6.4	3.2	77	9.8	41.4
	g- t	154	6.2	40	3.5	26.0	37	8.1	24.0	10	10.6	6.5	67	8.5	43.5
	g- g+	76	3.0	30	2.6	39.5	2	0.4	2.6	-	-	-	44	5.6	57.9
	g- g-	12	0.5	2	0.2	16.7	4	0.9	33.3	-	-	-	6	0.8	50.0
	Other	39	1.6	23	2.0	59.0	4	0.9	10.3	1	1.1	2.6	11	1.4	28.2
His	t t	30	3.1	9	3.2	30.0	13	5.4	43.3	-	-	-	8	2.0	26.7
	t -90	127	13.1	44	15.4	34.7	38	15.6	29.9	2	5.6	1.6	43	10.7	33.9
	t 90	161	16.7	76	26.6	47.2	39	16.1	24.2	5	13.9	3.1	41	10.2	25.5
	g+ t	49	5.1	17	5.9	34.7	14	5.8	28.6	1	2.8	2.0	17	4.2	34.7
	g+ -90	271	28.0	74	25.9	27.3	72	29.6	26.6	9	25.0	3.3	116	28.9	42.8
	g+ 90	175	18.1	50	17.5	28.6	26	10.7	14.9	5	13.9	2.9	94	23.4	53.7
	g- t	2	0.2	-	-	-	2	0.8	100.0	-	-	-	-	-	-
	g- -90	75	7.8	7	2.5	9.3	22	9.1	29.3	6	16.7	8.0	40	10.0	53.3
	g- 90	63	6.5	5	1.8	7.9	15	6.2	23.8	6	16.7	9.5	37	9.2	58.7
	Other	14	1.5	4	1.4	28.6	2	0.8	14.3	2	5.6	14.3	6	1.5	42.9

ISBN:978-84-694-2163-5/DL:T.885-2011

Appendix Table 1. *Continue (1).*

Ile	t	t		164	6.7	27	3.3	16.5	79	8.1	48.2	2	7.1	1.2	56	9.1	34.2
	t	g+		3	0.1	-	-	-	2	0.2	66.7	-	-	-	1	0.2	33.3
	t	g-		64	2.6	37	4.5	57.8	13	1.3	20.3	2	7.1	3.1	12	2.0	18.8
	g+	t		1425	58.2	593	71.9	41.6	564	57.6	39.6	11	39.3	0.8	257	41.8	18.0
	g+	g+		386	15.8	114	13.8	29.5	154	15.7	39.9	2	7.1	0.5	116	18.9	30.1
	g+	g-		44	1.8	8	1.0	18.2	21	2.1	47.7	1	3.6	2.3	14	2.3	31.8
	g-	t		270	11.0	25	3.0	9.3	109	11.1	40.4	7	25.0	2.6	129	21.0	47.8
	g-	g-		13	0.5	3	0.4	23.1	3	0.3	23.1	-	-	-	7	1.1	53.9
	Other			79	3.2	18	2.2	22.8	35	3.6	44.3	3	10.7	3.8	23	3.7	29.1
Leu	t	t		130	3.6	62	4.0	47.7	52	5.2	40.0	1	1.3	0.8	15	1.5	11.5
	t	g+		20	0.6	8	0.5	40.0	11	1.1	55.0	-	-	-	1	0.1	5.0
	t	g-		1008	27.9	460	29.4	45.6	372	37.1	36.9	13	17.3	1.3	163	16.7	16.2
	g+	t		1995	55.1	875	56.0	43.9	416	41.5	20.9	52	69.3	2.6	652	66.7	32.7
	g+	g+		49	1.4	18	1.2	36.7	16	1.6	32.7	2	2.7	4.1	13	1.3	26.5
	g+	g-		287	7.9	97	6.2	33.8	90	9.0	31.4	5	6.7	1.7	95	9.7	33.1
	g-	t		20	0.6	-	-	-	14	1.4	70.0	-	-	-	6	0.6	30.0
	g-	g-		29	0.8	3	0.2	10.3	15	1.5	51.7	1	1.3	3.5	10	1.0	34.5
	Other			81	2.2	41	2.6	50.6	16	1.6	19.8	1	1.3	1.2	23	2.4	28.4
Lys	t	t	t	448	17.6	212	21.0	47.3	118	24.4	26.3	11	9.7	2.5	107	11.4	23.9
	t	t	g+	88	3.5	38	3.8	43.2	28	5.8	31.8	2	1.8	2.3	20	2.1	22.7
	t	t	g-	105	4.1	57	5.7	54.3	26	5.4	24.8	-	-	-	22	2.3	21.0
	t	g+	t	42	1.7	21	2.1	50.0	4	0.8	9.5	3	2.7	7.1	14	1.5	33.3
	t	g+	g+	11	0.4	7	0.7	63.6	1	0.2	9.1	-	-	-	3	0.3	27.3
	t	g+	g-	1	0.0	1	0.1	100.0	-	-	-	-	-	-	-	-	-
	t	g-	t	93	3.7	54	5.4	58.1	17	3.5	18.3	5	4.4	5.4	17	1.8	18.3
	t	g-	g+	6	0.2	4	0.4	66.7	2	0.4	33.3	-	-	-	-	-	-
	t	g-	g-	21	0.8	10	1.0	47.6	4	0.8	19.1	3	2.7	14.3	4	0.4	19.1
	g+	t	t	762	29.9	292	29.0	38.3	129	26.7	16.9	44	38.9	5.8	297	31.5	39.0
	g+	t	g+	142	5.6	56	5.6	39.4	19	3.9	13.4	4	3.5	2.8	63	6.7	44.4
	g+	t	g-	154	6.0	61	6.1	39.6	26	5.4	16.9	3	2.7	2.0	64	6.8	41.6
	g+	g+	t	258	10.1	73	7.2	28.3	34	7.0	13.2	16	14.2	6.2	135	14.3	52.3
	g+	g+	g+	52	2.0	17	1.7	32.7	9	1.9	17.3	3	2.7	5.8	23	2.4	44.2
	g+	g+	g-	19	0.8	5	0.5	26.3	4	0.8	21.1	-	-	-	10	1.1	52.6
	g+	g-	t	59	2.3	22	2.2	37.3	10	2.1	17.0	2	1.8	3.4	25	2.7	42.4
	g+	g-	g+	4	0.2	1	0.1	25.0	1	0.2	25.0	-	-	-	2	0.2	50.0
	g+	g-	g-	9	0.4	-	-	-	1	0.2	11.1	1	0.9	11.1	7	0.7	77.8
	g-	t	t	101	4.0	14	1.4	13.9	22	4.6	21.8	4	3.5	4.0	61	6.5	60.4
	g-	t	g+	24	0.9	4	0.4	16.7	6	1.2	25.0	2	1.8	8.3	12	1.3	50.0
	g-	t	g-	20	0.8	10	1.0	50.0	5	1.0	25.0	1	0.9	5.0	4	0.4	20.0
	g-	g+	t	3	0.1	-	-	-	-	-	-	-	-	-	3	0.3	100.0
	g-	g-	t	4	0.2	2	0.2	50.0	-	-	-	-	-	-	2	0.2	50.0
	Other			122	4.8	47	4.7	38.5	18	3.7	14.8	9	8.0	7.4	48	5.1	39.3
Met	t	t	t	32	3.7	12	3.1	37.5	11	4.7	34.4	-	-	-	9	4.0	28.1
	t	t	g+	56	6.5	19	4.9	33.9	28	11.9	50.0	2	11.8	3.6	7	3.1	12.5
	t	t	g-	61	7.0	26	6.7	42.6	26	11.1	42.6	1	5.9	1.6	8	3.5	13.1
	t	g+	t	2	0.2	-	-	-	2	0.9	100.0	-	-	-	-	-	-
	t	g+	g+	17	2.0	5	1.3	29.4	11	4.7	64.7	-	-	-	1	0.4	5.9
	t	g-	t	20	2.3	8	2.1	40.0	7	3.0	35.0	1	5.9	5.0	4	1.8	20.0
	t	g-	g+	10	1.2	7	1.8	70.0	2	0.9	20.0	-	-	-	1	0.4	10.0
	t	g-	g-	50	5.8	27	7.0	54.0	7	3.0	14.0	1	5.9	2.0	15	6.6	30.0
	g+	t	t	75	8.7	39	10.1	52.0	15	6.4	20.0	1	5.9	1.3	20	8.9	26.7
	g+	t	g+	97	11.2	45	11.6	46.4	28	11.9	28.9	-	-	-	24	10.6	24.7
	g+	t	g-	142	16.4	68	17.5	47.9	28	11.9	19.7	4	23.5	2.8	42	18.6	29.6
	g+	g+	t	28	3.2	16	4.1	57.1	4	1.7	14.3	1	5.9	3.6	7	3.1	25.0
	g+	g+	g+	129	14.9	75	19.3	58.1	24	10.2	18.6	2	11.8	1.6	28	12.4	21.7
	g+	g+	g-	36	4.2	21	5.4	58.3	1	0.4	2.8	-	-	-	14	6.2	38.9
	g+	g-	t	1	0.1	-	-	-	1	0.4	100.0	-	-	-	-	-	-
	g+	g-	g+	1	0.1	1	0.3	100.0	-	-	-	-	-	-	-	-	-
	g+	g-	g-	4	0.5	-	-	-	1	0.4	25.0	-	-	-	3	1.3	75.0
	g-	t	t	19	2.2	1	0.3	5.3	9	3.8	47.4	1	5.9	5.3	8	3.5	42.1
	g-	t	g+	30	3.5	2	0.5	6.7	14	6.0	46.7	2	11.8	6.7	12	5.3	40.0
	g-	t	g-	23	2.7	5	1.3	21.7	9	3.8	39.1	-	-	-	9	4.0	39.1
	g-	g+	t	2	0.2	1	0.3	50.0	-	-	-	-	-	-	1	0.4	50.0
	g-	g+	g+	4	0.5	2	0.5	50.0	-	-	-	-	-	-	2	0.9	50.0
	g-	g-	t	2	0.2	-	-	-	2	0.9	100.0	-	-	-	-	-	-
	g-	g-	g+	1	0.1	-	-	-	-	-	-	1	5.9	100.0	-	-	-
	g-	g-	g-	2	0.2	-	-	-	-	-	-	-	-	-	2	0.9	100.0
	Other			22	2.5	8	2.1	36.4	5	2.1	22.7	-	-	-	9	4.0	40.9
Phe	t	90		623	34.9	332	53.6	53.3	119	22.1	19.1	15	31.3	2.4	157	27.1	25.2
	g+	90		946	53.0	269	43.4	28.4	299	55.6	31.6	27	56.3	2.9	351	60.5	37.1
	g-	90		212	11.9	18	2.9	8.5	119	22.1	56.1	5	10.4	2.4	70	12.1	33.0
	Other			5	0.3	1	0.2	20.0	1	0.2	20.0	1	2.1	20.0	2	0.3	40.0

ISBN:978-84-694-2163-5/DL:T.885-2011

Appendix Table 1. *Continue (2).*

Ser	t	606	23.0	135	18.4	22.3	164	31.3	27.1	29	18.5	4.8	278	22.8	45.9
	g+	816	31.0	337	45.9	41.3	153	29.2	18.8	30	19.1	3.7	296	24.3	36.3
	g-	1195	45.4	261	35.5	21.8	205	39.1	17.2	95	60.5	8.0	634	52.1	53.1
	Other	17	0.7	2	0.3	11.8	2	0.4	11.8	3	1.9	17.7	10	0.8	58.8
Thr	t	196	7.5	4	0.6	2.0	86	11.8	43.9	7	6.5	3.6	99	8.8	50.5
	g+	1194	45.6	462	70.6	38.7	398	54.7	33.3	28	25.9	2.4	306	27.1	25.6
	g-	1209	46.2	185	28.3	15.3	237	32.6	19.6	71	65.7	5.9	716	63.4	59.2
	Other	20	0.8	3	0.5	15.0	7	1.0	35.0	2	1.9	10.0	8	0.7	40.0
Trp	t -90	112	16.9	55	24.8	49.1	26	13.3	23.2	3	16.7	2.7	28	12.3	25.0
	t 0	36	5.4	9	4.1	25.0	13	6.6	36.1	1	5.6	2.8	13	5.7	36.1
	t 90	91	13.7	58	26.1	63.7	20	10.2	22.0	-	-	-	13	5.7	14.3
	g+ -90	29	4.4	2	0.9	6.9	17	8.7	58.6	1	5.6	3.5	9	4.0	31.0
	g+ 0	71	10.7	30	13.5	42.3	12	6.1	16.9	2	11.1	2.8	27	11.9	38.0
	g+ 90	227	34.2	55	24.8	24.2	78	39.8	34.4	7	38.9	3.1	87	38.3	38.3
	g- -90	47	7.1	9	4.1	19.2	13	6.6	27.7	-	-	-	25	11.0	53.2
	g- 0	5	0.8	-	-	-	-	-	-	-	-	-	5	2.2	100.0
	g- 90	41	6.2	3	1.4	7.3	17	8.7	41.5	4	22.2	9.8	17	7.5	41.5
	Other	4	0.6	1	0.5	25.0	-	-	-	-	-	-	3	1.3	75.0
Tyr	t 90	571	34.8	276	54.6	48.3	134	24.9	23.5	13	24.5	2.3	148	27.3	25.9
	g+ 90	856	52.2	212	41.9	24.8	294	54.7	34.4	34	64.2	4.0	316	58.2	36.9
	g- 90	209	12.7	17	3.4	8.1	110	20.5	52.6	5	9.4	2.4	77	14.2	36.8
	Other	4	0.2	1	0.2	25.0	-	-	-	1	1.9	25.0	2	0.4	50.0
Val	t	2244	71.3	759	83.9	33.8	980	72.3	43.7	35	58.3	1.6	470	57.0	20.9
	g+	572	18.2	80	8.8	14.0	224	16.5	39.2	12	20.0	2.1	256	31.0	44.8
	g-	237	7.5	45	5.0	19.0	114	8.4	48.1	10	16.7	4.2	68	8.2	28.7
	Other	93	3.0	21	2.3	22.6	38	2.8	40.9	3	5.0	3.2	31	3.8	33.3

* The table shows how the residues in the 1996 global, helical, β-strand, turn and coil samples are distributed among rotamers. The data for each rotamer in the various secondary structures are compiled in three columns. The first column shows the number of residues for each rotamer in each structure. The second shows the proportion of residues for the corresponding amino acid belonging to each rotamer in each secondary structure. The third shows how the residues for a single rotamer are distributed among helices, β-strands, turns and coils. The results in the table were obtained by using, when possible, combinations of side-chain dihedral angles from χ_1 to χ_3 in nondependent (Averaged) and dependent (Helix, β-Strand, Turn and Coil) secondary-structure coordinates. In the case of "Averaged", only the number of residues for the rotamer and the proportion throughout the amino acid sample are shown. • The χ_2 data from Asn and Asp and the χ_3 data from Gln and Glu were not used to build the "full rotamer libraries" because their dihedral angle distributions do not have well separated peaks.

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8: RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

Appendix Table 2. Mean χ_1 values and their standard deviations for dependent and nondependent secondary-structure "full rotamer libraries".

Residue	*Rotamer label* •	*Mean values and standard deviation for* χ_1									
		Averaged		*Helix*		*ß-strand*		*Turn*		*Coil*	
Arg	**t t t**	-176.1	14.4	-177.6	12.8	-177.9	16.7	-179.6	5.5	-171.8	14.1
	t t g+	-174.1	14.6	-173.8	14.5	-175.5	14.3	-	-	-172.9	15.4
	t t g-	-174.4	12.6	-174.5	12.5	-175.0	11.5	177.9	14.6	-172.7	13.4
	t g+ t	-165.5	13.8	-166.1	15.7	-	-	-	-	-164.5	10.0
	t g+ g+	-153.6	21.1	-153.0	20.8	-177.1	0.0	-	-	-133.7	0.0
	t g+ g-	-155.8	0.0	-155.8	0.0	-	-	-	-	-	-
	t g- t	-178.5	11.9	-177.8	11.8	-176.6	12.1	170.3	7.3	-179.9	11.7
	t g- g+	-177.1	13.5	-179.8	11.2	-152.0	0.0	-	-	175.7	1.8
	t g- g-	177.5	13.7	177.0	14.0	178.4	14.8	-178.4	0.0	-179.2	0.0
	g+ t t	-67.7	12.6	-69.0	12.4	-66.4	13.2	-60.7	6.8	-66.9	12.6
	g+ t g+	-68.0	14.3	-71.5	14.0	-65.2	16.1	-59.6	15.7	-66.0	13.0
	g+ t g-	-68.6	16.4	-72.0	14.9	-62.4	10.6	-62.9	15.8	-68.5	18.3
	g+ g+ t	-64.0	13.6	-68.0	10.6	-63.2	12.6	-81.4	24.8	-61.4	12.8
	g+ g+ g+	-63.8	15.4	-63.7	11.6	-63.8	23.2	-72.4	10.6	-62.4	13.0
	g+ g+ g-	-59.4	15.3	-72.2	21.5	-63.5	4.5	-46.9	0.0	-53.5	7.0
	g+ g- t	-88.7	12.3	-89.8	3.7	-95.3	15.6	-	-	-87.0	12.6
	g+ g- g+	-117.0	6.6	-118.6	6.8	-119.9	0.0	-	-	-109.2	0.0
	g+ g- g-	-85.8	17.4	-89.4	0.0	-79.7	16.1	-	-	-91.2	18.4
	g- t t	65.3	11.5	73.9	10.2	62.1	10.7	75.0	10.0	63.4	10.6
	g- t g+	62.3	13.5	66.7	14.5	55.6	10.1	63.5	1.8	62.1	13.7
	g- t g-	65.4	13.8	74.7	18.0	64.4	12.6	-	-	63.0	11.4
	g- g+ t	103.4	1.1	104.5	0.0	-	-	-	-	102.4	0.0
	g- g+ g+	49.9	1.7	-	-	-	-	-	-	49.9	1.7
	g- g+ g-	74.3	0.0	-	-	-	-	-	-	74.3	0.0
	g- g- t	52.4	10.5	48.0	0.0	43.1	0.0	-	-	55.8	11.3
Asn	**t**	-170.7	12.9	-172.7	12.0	-172.9	13.4	-169.3	13.8	-169.4	12.6
	g+	-70.3	13.0	-73.2	10.7	-68.3	12.5	-69.6	11.8	-68.5	14.4
	g-	63.1	10.8	63.7	14.0	64.0	9.3	63.5	7.9	62.8	10.7
Asp	**t**	-171.3	12.3	-170.2	11.9	-172.1	12.5	-169.5	11.7	-171.6	12.3
	g+	-69.9	11.3	-71.9	10.0	-69.2	9.9	-67.1	11.6	-68.3	12.7
	g-	61.0	11.2	57.8	14.0	63.3	10.2	61.8	10.3	61.0	10.8
Cys	**t**	-176.4	12.3	-175.5	12.7	-176.9	10.7	-178.0	13.3	-176.2	13.1
	g+	-64.9	10.3	-68.2	9.6	-63.0	10.6	-61.9	8.6	-63.5	10.3
	g-	64.1	11.2	62.6	11.5	64.7	7.2	68.0	15.2	63.5	11.6
Gln	**t t**	-175.3	14.0	-174.6	16.6	-177.1	11.3	-174.9	8.0	-175.2	11.2
	t g+	-168.6	20.2	-170.0	20.2	-163.8	24.2	-	-	-167.7	17.9
	t g-	-175.1	13.2	-174.2	12.0	-179.9	10.3	177.4	2.8	-173.9	17.9
	g+ t	-67.7	12.7	-70.1	10.3	-63.6	12.8	-68.4	15.1	-65.8	14.8
	g+ g+	-65.4	14.6	-69.3	13.0	-64.2	16.0	-64.6	21.0	-62.6	14.0
	g+ g-	-71.6	21.7	-74.5	17.8	-65.2	24.1	-77.4	19.9	-69.2	23.4
	g- t	63.6	14.7	60.2	12.4	62.0	15.4	57.4	17.2	66.5	13.9
	g- g+	65.3	17.9	56.8	19.8	62.9	23.3	-	-	70.1	11.3
	g- g-	62.3	18.6	73.8	2.9	35.4	13.9	80.6	13.0	64.1	7.6
Glu	**t t**	-175.7	13.1	-175.3	14.2	-177.0	11.7	-178.0	13.4	-174.0	12.2
	t g+	-165.8	19.4	-166.9	18.8	-162.8	18.4	-167.3	0.0	-165.2	20.9
	t g-	-170.9	16.4	-171.2	16.1	-177.9	15.7	-170.0	13.9	-166.3	16.0
	g+ t	-67.9	14.5	-69.5	11.8	-66.9	15.4	-67.6	12.9	-66.3	17.2
	g+ g+	-69.1	16.4	-71.4	16.5	-64.8	13.4	-69.9	12.8	-66.4	16.9
	g+ g-	-65.6	20.0	-68.2	18.7	-76.4	29.2	-76.6	20.8	-60.8	19.6
	g- t	63.0	16.9	69.8	17.6	60.5	14.4	62.7	19.3	60.4	16.4
	g- g+	56.9	19.0	48.6	19.0	58.1	20.9	-	-	62.6	16.6
	g- g-	48.3	25.7	63.1	12.4	44.1	13.1	-	-	46.1	32.6

Appendix Table 2. *Continue (1).*

His	t	t		-173.9	11.6	-168.4	10.0	-177.9	7.7	-	-	-173.5	15.4
	t	-90		-172.8	13.3	-173.7	13.1	-178.0	13.7	-173.1	7.5	-167.4	11.4
	t	90		-178.5	12.0	179.9	11.8	-179.9	10.7	-177.1	9.9	-174.4	12.8
	g+	t		-66.9	9.3	-71.8	5.6	-66.8	5.6	-45.0	0.0	-63.3	11.3
	g+	-90		-64.5	11.9	-69.1	12.1	-64.1	11.6	-61.0	8.4	-62.2	11.4
	g+	90		-67.8	11.6	-72.5	12.4	-65.3	11.5	-55.9	10.7	-66.7	10.3
	g-	t		62.1	1.1	-	-	62.1	1.1	-	-	-	-
	g-	-90		61.4	10.8	73.3	9.3	60.0	11.1	66.1	4.0	59.4	10.0
	g-	90		59.3	9.6	62.0	6.5	63.9	8.0	61.0	9.6	56.8	9.7
Ile	t	t		-171.3	9.8	-167.4	9.3	-175.0	8.6	-164.5	10.9	-168.1	9.6
	t	g+		-163.9	14.9	-	-	-171.0	13.5	-	-	-149.6	0.0
	t	g-		-166.8	9.5	-166.1	9.7	-173.3	9.2	-161.8	6.5	-162.8	5.5
	g+	t		-64.4	8.0	-67.8	6.5	-61.5	7.8	-65.2	6.1	-62.9	8.4
	g+	g+		-58.3	8.4	-62.0	6.9	-56.2	8.9	-62.4	5.4	-57.3	7.9
	g+	g-		-76.3	13.6	-89.8	8.0	-69.5	12.7	-73.5	0.0	-78.9	11.0
	g-	t		61.0	8.0	63.0	10.0	61.0	7.2	56.1	8.4	60.8	8.0
	g-	g-		51.7	12.6	56.3	6.9	56.3	11.1	-	-	47.8	13.7
Leu	t	t		-158.9	15.7	-156.7	16.2	-164.6	13.4	-169.8	0.0	-147.2	12.5
	t	g+		-164.9	15.3	-163.7	12.2	-168.6	14.8	-	-	-134.6	0.0
	t	g-		-177.9	12.3	-177.5	11.7	179.6	11.6	-171.7	8.0	-173.7	14.0
	g+	t		-65.5	10.4	-69.0	9.8	-60.9	10.1	-65.5	7.2	-63.8	9.9
	g+	g+		-84.4	14.4	-89.0	17.6	-86.5	11.6	-71.9	0.8	-77.6	8.8
	g+	g-		-93.4	15.2	-101.2	10.8	-84.6	17.3	-93.8	8.1	-93.8	12.6
	g-	t		72.5	10.3	-	-	70.8	10.3	-	-	76.4	9.3
	g-	g-		63.6	15.6	76.7	7.3	58.7	14.4	54.9	0.0	68.0	16.4
Lys	t	t	t	-175.3	13.4	-177.1	13.9	-174.8	11.6	179.2	11.2	-171.7	13.6
	t	t	g+	-172.2	18.2	-171.8	17.5	-177.2	16.2	-159.9	12.3	-167.3	20.4
	t	t	g-	-179.3	17.9	178.4	17.3	-176.0	7.8	-	-	-177.1	25.5
	t	g+	t	-169.7	25.5	-169.2	25.8	-166.4	15.4	-166.5	27.5	-172.0	26.7
	t	g+	g+	-171.0	15.9	-174.9	17.9	-156.9	0.0	-	-	-166.6	7.5
	t	g+	g-	-178.7	0.0	-178.7	0.0	-	-	-	-	-	-
	t	g-	t	-178.5	15.8	-179.8	16.9	-178.1	11.8	-177.2	8.5	-175.3	16.5
	t	g-	g+	-177.7	18.3	-173.7	20.4	174.3	8.4	-	-	-	-
	t	g-	g-	-172.6	18.4	-174.3	15.1	176.9	22.1	-157.5	12.2	-169.0	17.4
	g+	t	t	-68.8	13.0	-71.8	13.2	-68.1	13.7	-62.0	12.4	-67.1	11.7
	g+	t	g+	-67.9	13.4	-69.8	9.6	-65.2	10.3	-50.9	22.3	-68.2	15.3
	g+	t	g-	-70.2	16.9	-73.2	14.5	-67.6	14.5	-71.1	9.5	-68.4	19.5
	g+	g+	t	-62.9	14.9	-66.9	11.9	-62.7	15.5	-58.5	9.4	-61.3	16.3
	g+	g+	g+	-58.4	15.8	-61.8	18.4	-58.3	15.1	-45.8	6.4	-57.5	13.7
	g+	g+	g-	-60.7	17.2	-59.6	5.2	-69.7	29.0	-	-	-57.7	13.0
	g+	g-	t	-85.2	21.1	-91.1	14.4	-70.8	22.3	-87.6	12.0	-85.6	23.4
	g+	g-	g+	-90.9	10.3	-108.0	0.0	-90.0	0.0	-	-	-82.7	0.4
	g+	g-	g-	-89.2	23.0	-	-	-79.2	0.0	-58.8	0.0	-95.0	22.3
	g-	t	t	66.1	12.8	73.3	12.3	62.1	10.4	70.0	8.5	65.6	13.2
	g-	t	g+	68.6	15.1	74.8	17.8	60.7	6.8	84.1	22.1	67.9	12.8
	g-	t	g-	65.6	10.0	64.6	11.3	67.3	10.1	61.6	0.0	67.0	6.7
	g-	g+	t	64.6	5.6	-	-	-	-	-	-	64.6	5.6
	g-	g-	t	43.9	5.7	41.4	7.1	-	-	-	-	46.5	1.0
Met	t	t	t	-175.5	10.0	-176.6	9.4	-175.6	7.6	-	-	-173.9	12.7
	t	t	g+	-175.4	12.5	-176.1	14.1	-175.9	8.1	171.5	16.5	-167.8	16.2
	t	t	g-	-177.1	10.1	-178.9	10.1	-176.8	10.2	-165.2	0.0	-174.1	8.5
	t	g+	t	-163.9	4.9	-	-	-163.9	4.9	-	-	-	-
	t	g+	g+	-173.8	8.7	-171.7	10.3	-174.0	7.8	-	-	177.5	0.0
	t	g-	t	-173.1	8.0	-171.4	6.8	-172.8	10.1	-172.8	0.0	-177.2	5.5
	t	g-	g+	-174.6	22.7	-171.1	26.3	176.8	2.6	-	-	178.5	0.0
	t	g-	g-	-172.8	11.9	-172.5	13.4	-177.5	7.5	178.7	0.0	-170.7	10.2

ISBN:978-84-694-2163-5/DL:T.885-2011

Appendix Table 2. *Continue (2).*

	g+	t	t	-67.5	9.5	-71.1	9.4	-62.7	6.7	-68.2	0.0	-64.0	8.8
	g+	t	g+	-67.8	8.8	-69.1	7.6	-67.6	6.7	-	-	-65.7	12.0
	g+	t	g-	-70.2	11.7	-72.4	10.0	-67.7	14.4	-69.1	18.5	-68.2	10.7
	g+	g+	t	-66.5	13.5	-68.8	16.3	-64.1	9.6	-68.1	0.0	-62.3	6.4
	g+	g+	g+	-67.4	10.8	-70.9	9.8	-60.1	9.3	-58.2	6.8	-65.1	10.9
	g+	g+	g-	-65.1	10.7	-68.0	11.5	-59.3	0.0	-	-	-61.2	8.3
	g+	g-	t	-71.9	0.0	-	-	-71.9	0.0	-	-	-	-
	g+	g-	g+	-75.2	0.0	-75.2	0.0	-	-	-	-	-	-
	g+	g-	g-	-104.3	11.0	-	-	-116.3	0.0	-	-	-100.3	9.9
	g-	t	t	57.2	15.3	70.9	0.0	59.4	15.2	62.1	0.0	52.3	15.5
	g-	t	g+	63.8	12.1	53.3	11.1	62.6	12.8	67.3	2.2	66.3	11.2
	g-	t	g-	62.5	12.8	73.2	13.8	58.7	9.4	-	-	60.4	11.9
	g-	g+	t	60.5	19.3	41.2	0.0	-	-	-	-	79.8	0.0
	g-	g+	g+	68.1	9.8	74.2	10.7	-	-	-	-	61.9	1.4
	g-	g-	t	60.3	6.7	-	-	60.3	6.7	-	-	-	-
	g-	g-	g+	47.6	0.0	-	-	-	-	47.6	0.0	-	-
	g-	g-	g-	40.7	7.1	-	-	-	-	-	-	40.7	7.1
Phe	t	90		-178.0	10.5	-179.3	10.8	-178.1	9.5	-172.6	8.5	-175.5	10.2
	g+	90		-67.1	11.2	-71.7	12.3	-65.6	9.6	-57.7	9.0	-65.5	10.6
	g-	90		63.2	10.0	67.8	10.3	62.9	9.2	63.2	12.0	62.4	10.8
Ser	t			-179.3	13.2	-178.1	13.1	-178.3	11.0	-177.2	13.2	179.4	14.3
	g+			-64.7	14.4	-64.8	12.1	-64.3	12.6	-65.2	18.6	-64.8	16.8
	g-			64.1	12.9	65.3	12.8	62.3	11.6	64.3	14.0	64.2	13.1
Thr	t			-172.0	10.3	-157.4	9.5	-171.5	10.0	-170.3	7.6	-173.1	10.4
	g+			-60.0	11.3	-60.4	10.1	-60.2	9.6	-60.5	11.3	-59.3	14.5
	g-			60.9	10.2	61.9	10.6	60.7	10.1	60.9	9.5	60.7	10.3
Trp	t	-90		-176.1	10.8	-177.2	10.2	-179.5	11.1	-168.1	2.2	-171.5	10.0
	t	0		-172.9	10.6	-171.6	6.6	-172.0	11.1	-173.1	0.0	-174.7	12.3
	t	90		178.3	8.7	179.5	8.9	174.1	6.9	-	-	179.4	8.2
	g+	-90		-65.0	10.2	-70.6	8.5	-69.5	6.2	-55.1	0.0	-56.4	10.9
	g+	0		-68.8	9.1	-70.4	9.7	-67.2	8.1	-71.6	6.4	-67.4	8.8
	g+	90		-68.5	10.9	-75.4	11.5	-67.4	8.9	-64.8	11.8	-65.5	10.1
	g-	-90		62.7	10.7	70.2	8.7	60.8	12.6	-	-	61.0	8.9
	g-	0		72.8	5.4	-	-	-	-	-	-	72.8	5.4
	g-	90		60.8	12.1	76.5	18.3	59.7	13.4	63.0	9.5	58.8	6.5
Tyr	t	90		-179.0	11.0	179.8	11.0	-179.2	10.9	180.0	6.1	-176.7	11.0
	g+	90		-65.6	10.9	-69.5	10.6	-65.0	10.3	-61.4	7.5	-63.9	11.2
	g-	90		63.8	10.9	72.5	10.6	64.7	9.1	64.6	14.8	60.6	11.7
Val	t			175.7	8.4	172.2	7.1	178.1	7.9	175.6	10.1	176.4	9.2
	g+			-62.4	10.8	-58.7	13.9	-62.2	9.4	-62.1	11.2	-63.7	10.5
	g-			65.0	11.7	70.2	10.4	61.2	10.3	64.7	7.1	67.7	13.1

The table shows the mean χ_1 values (first column) and their corresponding standard deviations (second column) for the different rotamers in the 1996 "full rotamer libraries" (except for rotamers labelled as "Other"). The data in the table were obtained by using, when possible, combinations of side-chain dihedral angles from χ_1 to χ_3 in nondependent (Averaged) and dependent (Helix, β-Strand, Turn and Coil) secondary-structure coordinates. Residues grouped as "Other" (see **Appendix Table 1**) were not taken into account because they are not a homogeneous sample. • The χ_2 data from Asn and Asp and the χ_3 data from Gln and Glu were not used to build the "full rotamer libraries" because their dihedral angle distributions do not have well separated peaks.

ISBN:978-84-694-2163-5/DL:T.885-2011

Appendix Table 3. Mean χ_2 values and their standard deviations for dependent and nondependent secondary-structure "full rotamer libraries".

Residue	*Rotamer label* •			*Mean values and standard deviation for* χ_2									
				Averaged		*Helix*		*ß-strand*		*Turn*		*Coil*	
Arg	t	t	t	178.1	20.3	177.2	20.8	177.0	21.1	180.0	6.2	-179.8	19.4
	t	t	g+	-176.9	22.5	-173.7	25.1	-179.4	21.9	-	-	178.6	13.8
	t	t	g-	178.9	16.3	-178.5	17.7	172.6	10.1	161.3	6.5	176.5	12.8
	t	g+	t	-82.9	23.9	-91.2	21.1	-	-	-	-	-69.1	21.7
	t	g+	g+	-80.5	22.5	-77.1	22.8	-72.6	0.0	-	-	-108.8	0.0
	t	g+	g-	-55.2	0.0	-55.2	0.0	-	-	-	-	-	-
	t	g-	t	72.4	15.2	69.7	14.1	78.7	10.2	68.4	10.8	78.8	17.9
	t	g-	g+	83.2	14.0	74.0	5.9	105.8	0.0	-	-	90.3	11.6
	t	g-	g-	66.0	12.0	65.0	11.8	72.8	12.5	65.0	0.0	56.7	0.0
	g+	t	t	180.0	15.7	179.3	14.0	-177.9	18.8	177.9	9.3	179.9	16.4
	g+	t	g+	-173.1	18.2	-169.0	13.1	-176.9	21.9	175.0	17.7	-175.2	20.3
	g+	t	g-	178.4	20.7	177.0	19.9	179.5	15.5	-175.4	10.3	178.6	23.3
	g+	g+	t	-70.8	12.1	-70.8	12.6	-70.3	8.8	-79.4	15.3	-70.6	13.7
	g+	g+	g+	-67.6	14.2	-65.9	13.6	-72.3	12.6	-60.4	14.7	-67.7	14.5
	g+	g+	g-	-81.2	13.4	-79.1	10.5	-70.1	2.5	-91.2	0.0	-83.4	14.9
	g+	g-	t	69.4	18.8	69.9	14.2	60.7	6.6	-	-	71.1	21.1
	g+	g-	g+	75.1	13.7	83.8	9.0	70.0	0.0	-	-	54.0	0.0
	g+	g-	g-	91.0	16.5	80.6	0.0	95.8	15.3	-	-	88.2	17.7
	g-	t	t	179.4	19.1	-177.7	24.2	179.6	15.8	162.9	15.4	179.3	18.4
	g-	t	g+	-177.6	20.0	178.4	18.4	171.9	19.9	-178.5	6.2	-171.4	19.6
	g-	t	g-	179.6	19.6	173.9	8.2	-176.4	22.5	-	-	177.9	19.0
	g-	g+	t	-91.2	16.4	-107.6	0.0	-	-	-	-	-74.8	0.0
	g-	g+	g+	-73.5	1.9	-	-	-	-	-	-	-73.5	1.9
	g-	g+	g-	-86.9	0.0	-	-	-	-	-	-	-86.9	0.0
	g-	g-	t	87.5	8.6	85.7	0.0	89.3	0.0	-	-	87.5	10.4
Gln	t	t		-179.7	15.8	-178.4	16.0	178.9	11.0	-152.2	18.8	178.7	17.3
	t	g+		-87.1	17.7	-90.6	15.9	-74.8	15.8	-	-	-85.1	19.1
	t	g-		66.2	10.2	65.9	10.2	66.1	7.5	64.1	1.6	67.6	12.2
	g+	t		177.9	14.8	176.7	13.6	178.7	16.2	-179.2	13.6	179.2	15.9
	g+	g+		-65.9	15.6	-63.0	14.7	-71.5	14.7	-69.0	22.1	-66.5	15.3
	g+	g-		77.0	17.2	70.2	18.0	89.1	10.5	75.9	13.3	79.6	16.7
	g-	t		-179.1	14.7	-176.5	11.1	-179.8	10.1	-173.6	9.1	179.8	18.0
	g-	g+		-82.1	11.7	-95.3	7.8	-79.7	7.6	-	-	-77.2	10.0
	g-	g-		86.0	17.5	93.4	8.3	75.9	28.3	89.7	12.1	86.2	10.9
Glu	t	t		178.7	15.3	179.6	15.9	176.5	12.7	173.3	14.7	-179.0	17.4
	t	g+		-84.3	13.3	-82.0	13.2	-82.3	14.5	-98.6	0.0	-87.0	12.2
	t	g-		66.0	16.0	65.6	16.2	66.7	13.8	57.7	19.8	67.1	16.0
	g+	t		179.2	16.4	177.4	14.6	-178.6	14.1	-178.2	22.9	-179.7	18.6
	g+	g+		-66.8	16.5	-64.7	17.1	-77.5	15.1	-63.7	13.9	-67.6	15.0
	g+	g-		81.2	14.9	79.9	14.8	77.4	14.9	90.4	15.2	82.5	14.6
	g-	t		-177.7	17.9	-177.0	20.9	-178.4	13.5	176.4	13.1	-176.7	18.6
	g-	g+		-81.9	14.9	-79.5	16.8	-89.9	25.8	-	-	-83.2	12.4
	g-	g-		85.7	20.0	107.6	11.4	78.0	20.8	-	-	83.5	16.4
His	t	t		-179.2	15.8	179.6	18.2	-179.6	13.6	-	-	-177.1	16.0
	t	-90		-95.8	25.7	-100.5	18.1	-102.1	21.9	-84.9	41.3	-86.1	30.9
	t	90		78.9	20.2	77.4	17.4	76.9	20.1	81.1	13.5	83.3	24.7
	g+	t		172.4	15.1	175.9	15.0	169.8	13.5	153.2	0.0	172.4	15.5
	g+	-90		-75.3	22.9	-67.0	23.6	-84.8	23.6	-68.3	14.8	-75.3	20.0
	g+	90		96.5	23.7	98.2	24.9	90.6	22.1	111.6	13.3	96.4	23.5
	g-	t		-173.9	12.3	-	-	-173.9	12.3	-	-	-	-
	g-	-90		-84.6	18.7	-75.2	18.6	-92.5	18.0	-75.5	13.7	-83.2	18.0
	g-	90		89.1	16.4	89.5	27.1	88.6	12.9	88.6	14.4	89.3	16.1
Ile	t	t		167.3	11.3	170.1	11.8	165.8	10.7	175.3	1.5	167.9	11.5
	t	g+		-89.6	16.1	-	-	-99.3	10.4	-	-	-70.2	0.0
	t	g-		66.8	10.3	64.3	9.2	74.6	8.8	66.0	1.9	66.4	11.7
	g+	t		169.3	10.1	169.2	8.7	169.3	11.0	172.4	7.5	169.3	10.9
	g+	g+		-61.5	11.4	-59.5	11.2	-62.8	12.0	-56.3	6.3	-61.8	10.5
	g+	g-		74.6	16.7	55.0	7.4	80.9	13.0	69.3	0.0	76.7	17.6

ISBN:978-84-694-2163-5/DL:T.885-2011

Appendix Table 3. *Continue (1).*

	g-	t		170.4	11.2	171.1	11.8	171.5	10.8	171.3	6.2	169.3	11.5
	g-	g-		83.2	10.7	82.8	10.5	89.3	7.1	-	-	80.7	11.1
Leu	t	t		-176.8	27.6	-172.4	28.6	176.4	25.9	158.2	0.0	-169.8	24.5
	t	g+		-69.3	12.8	-76.4	11.2	-67.1	8.2	-	-	-36.5	0.0
	t	g-		63.9	11.0	62.1	10.0	66.3	11.1	63.1	10.7	63.4	12.4
	g+	t		175.4	10.0	175.3	9.6	175.2	10.7	176.7	10.7	175.6	10.0
	g+	g+		-57.0	11.4	-58.3	10.3	-62.9	8.6	-60.7	0.5	-47.4	10.6
	g+	g-		42.1	26.9	30.0	21.3	57.4	28.4	43.3	17.8	39.9	23.7
	g-	t		166.8	17.5	-	-	164.5	15.2	-	-	172.0	21.2
	g-	g-		78.8	16.4	80.2	1.0	78.6	14.7	52.3	0.0	81.5	19.4
Lys	t	t	t	179.3	17.8	178.5	16.8	178.4	19.3	178.6	15.3	-178.1	17.7
	t	t	g+	-175.8	20.9	-177.4	19.4	-176.4	21.1	170.7	13.4	-170.4	22.9
	t	t	g-	174.3	21.1	172.3	22.7	168.5	12.0	-	-	-173.8	20.4
	t	g+	t	-87.0	25.1	-85.7	27.2	-87.5	9.9	-72.3	40.2	-91.9	18.4
	t	g+	g+	-97.2	17.3	-106.9	9.6	-60.9	0.0	-	-	-86.6	10.6
	t	g+	g-	-110.4	0.0	-110.4	0.0	-	-	-	-	-	-
	t	g-	t	67.1	11.4	65.1	10.3	68.3	14.9	65.6	6.8	72.5	10.1
	t	g-	g+	87.4	12.1	89.7	11.3	82.7	12.2	-	-	-	-
	t	g-	g-	67.2	18.1	65.1	17.0	74.6	13.3	57.4	13.3	72.5	23.0
	g+	t	t	-179.4	17.0	-179.3	15.8	-178.0	16.1	-175.2	21.4	179.3	17.7
	g+	t	g+	-173.2	21.1	-171.5	14.4	-174.8	18.6	-155.4	16.3	-175.4	25.9
	g+	t	g-	175.3	21.7	173.1	19.5	174.4	24.5	-173.3	14.3	177.3	22.3
	g+	g+	t	-67.4	17.7	-66.4	16.4	-68.7	15.1	-72.7	12.7	-67.1	19.4
	g+	g+	g+	-70.1	23.7	-72.2	23.4	-77.4	19.9	-74.1	18.0	-65.2	24.8
	g+	g+	g-	-83.2	17.8	-75.4	26.4	-82.6	15.5	-	-	-87.3	10.6
	g+	g-	t	68.9	14.4	66.3	14.3	74.8	12.8	73.3	9.1	68.5	14.7
	g+	g-	g+	59.9	22.5	58.0	0.0	54.4	0.0	-	-	63.7	31.3
	g+	g-	g-	69.3	15.2	-	-	68.1	0.0	49.7	0.0	72.2	15.3
	g-	t	t	179.3	18.6	173.0	20.3	178.3	18.2	177.7	10.7	-178.9	18.4
	g-	t	g+	-168.8	20.5	-157.8	23.8	-168.3	13.8	-175.6	20.0	-171.6	20.9
	g-	t	g-	177.7	22.8	167.3	17.5	-178.7	26.5	171.5	0.0	-159.3	11.9
	g-	g+	t	-67.8	26.9	-	-	-	-	-	-	-67.8	26.9
	g-	g-	t	68.4	8.7	69.5	8.6	-	-	-	-	67.4	8.7
Met	t	t	t	175.9	12.1	176.4	12.8	177.4	11.9	-	-	173.4	10.7
	t	t	g+	-178.8	13.0	178.7	17.2	-178.1	8.1	177.6	3.9	-174.1	15.4
	t	t	g-	179.1	10.8	-176.7	9.9	176.3	10.6	-178.7	0.0	173.9	9.8
	t	g+	t	-96.9	19.9	-	-	-96.9	19.9	-	-	-	-
	t	g+	g+	-85.8	8.9	-84.7	11.5	-86.2	7.8	-	-	-86.9	0.0
	t	g-	t	65.1	16.2	59.2	17.0	70.9	17.5	59.4	0.0	68.3	7.1
	t	g-	g+	84.7	18.7	85.4	20.8	79.1	13.7	-	-	90.5	0.0
	t	g-	g-	65.8	14.4	65.9	16.0	68.6	14.6	60.2	0.0	64.5	11.0
	g+	t	t	178.8	12.0	176.5	10.5	-179.3	10.6	175.5	0.0	-178.1	14.5
	g+	t	g+	-178.3	12.6	-172.2	7.0	175.2	16.3	-	-	177.9	10.3
	g+	t	g-	178.6	11.5	176.4	12.2	-177.6	11.4	-178.2	11.9	179.3	9.4
	g+	g+	t	-66.5	13.7	-61.2	9.7	-66.1	9.1	-88.1	0.0	-75.7	16.6
	g+	g+	g+	-62.0	12.8	-60.2	12.7	-65.1	9.9	-54.0	0.6	-64.8	14.4
	g+	g+	g-	-67.9	15.7	-63.3	10.7	-63.9	0.0	-	-	-75.2	19.3
	g+	g-	t	79.9	0.0	-	-	79.9	0.0	-	-	-	-
	g+	g-	g+	64.7	0.0	64.7	0.0	-	-	-	-	-	-
	g+	g-	g-	89.9	29.3	-	-	60.2	0.0	-	-	99.8	27.5
	g-	t	t	-178.0	13.5	160.8	0.0	-174.8	13.5	178.7	0.0	-178.5	12.5
	g-	t	g+	-176.1	15.6	158.1	24.7	-176.2	12.0	-173.0	13.5	-172.2	13.3
	g-	t	g-	-175.1	13.4	171.9	14.2	-165.9	11.4	-	-	-177.2	5.7
	g-	g+	t	-80.4	27.9	-108.3	0.0	-	-	-	-	-52.4	0.0
	g-	g+	g+	-84.6	7.7	-86.0	5.9	-	-	-	-	-83.1	9.0
	g-	g-	t	92.8	13.5	-	-	92.8	13.5	-	-	-	-
	g-	g-	g+	102.5	0.0	-	-	-	-	102.5	0.0	-	-
	g-	g-	g-	67.3	7.2	-	-	-	-	-	-	67.3	7.2
Phe	t	90		75.3	18.3	76.4	17.8	75.1	20.8	69.4	23.3	73.7	16.4
	g+	90		101.7	30.0	106.4	36.7	94.4	24.6	114.6	24.3	103.3	27.4
	g-	90		90.6	11.7	85.0	17.1	92.7	9.1	84.6	8.8	89.0	13.2

ISBN:978-84-694-2163-5/DL:T.885-2011

Appendix Table 3. *Continue (2).*

Trp	t	**-90**	-103.5	15.3	-103.3	15.9	-99.7	17.5	-103.9	2.8	-107.2	11.2
	t	**0**	10.7	24.1	7.0	11.0	4.4	29.2	31.9	0.0	17.8	23.3
	t	**90**	78.9	13.2	80.4	12.0	76.0	15.5	-	-	76.4	13.7
	g+	**-90**	-90.6	10.2	-98.2	5.2	-93.0	10.2	-82.3	0.0	-85.2	8.5
	g+	**0**	-12.4	17.8	-15.7	18.1	-5.1	19.1	-22.0	16.2	-11.2	15.8
	g+	**90**	95.8	18.5	102.7	18.3	90.9	15.4	95.0	19.4	96.0	19.6
	g-	**-90**	-89.9	9.9	-91.4	7.7	-89.5	12.1	-	-	-89.5	9.2
	g-	**0**	-31.6	26.2	-	-	-	-	-	-	-31.6	26.2
	g-	**90**	88.1	13.5	87.7	21.6	87.6	12.4	89.4	11.8	88.4	13.0
Tyr	t	**90**	75.6	20.7	77.6	19.4	72.9	23.4	75.8	15.7	74.3	20.7
	g+	**90**	102.7	26.2	113.2	31.6	93.7	20.3	110.0	20.4	103.1	24.5
	g-	**90**	89.7	12.6	84.6	11.0	91.0	10.1	85.3	3.9	89.4	15.6

The table shows the mean χ_2 values (first column) and their corresponding standard deviations (second column) for the different rotamers in the 1996 "full rotamer libraries" (except for rotamers labelled as "Other"). The data in the table were obtained by using, when possible, combinations of side-chain dihedral angles from χ_1 to χ_3 in nondependent (Averaged) and dependent (Helix, β-Strand, Turn and Coil) secondary-structure coordinates. Residues grouped as "Other" (see **Appendix Table 1**) were not taken into account because they are not a homogeneous sample. • The χ_2 data from Asn and Asp and the χ_3 data from Gln and Glu were not used to build the "full rotamer libraries" because their dihedral angle distributions do not have well separated peaks. Consequently, results for Asn and Asp do not appear in this table.

Appendix Table 4. Mean χ_3 values and their standard deviations for dependent and nondependent secondary-structure "full rotamer libraries".

Residue	*Rotamer label* •			*Mean values and standard deviation for* χ_3									
				Averaged		*Helix*		*ß-strand*		*Turn*		*Coil*	
Arg	t	t	t	179.1	20.0	-178.4	20.5	174.7	20.9	-178.8	19.6	179.0	17.4
	t	t	g+	-65.9	17.3	-65.3	17.8	-66.0	16.5	-	-	-67.2	16.8
	t	t	g-	67.5	17.0	67.2	17.2	68.3	17.6	61.7	3.3	68.0	15.6
	t	g+	t	173.2	13.8	166.4	13.2	-	-	-	-	-175.6	4.1
	t	g+	g+	-67.7	16.7	-65.3	16.5	-60.1	0.0	-	-	-90.2	0.0
	t	g+	g-	104.7	0.0	104.7	0.0	-	-	-	-	-	-
	t	g-	t	-179.1	20.8	-177.6	16.6	173.4	16.9	176.8	5.9	-179.7	31.6
	t	g-	g+	-85.2	12.5	-88.9	15.4	-83.1	0.0	-	-	-79.0	0.1
	t	g-	g-	67.3	18.1	66.1	16.8	75.0	22.5	77.1	0.0	47.8	0.0
	g+	t	t	-178.8	18.2	-177.3	17.5	-177.7	19.2	178.5	7.9	178.5	18.8
	g+	t	g+	-71.3	20.9	-72.7	21.4	-70.7	22.3	-68.1	15.1	-70.2	20.1
	g+	t	g-	68.0	17.9	67.6	17.6	65.2	15.1	70.5	12.3	69.0	19.2
	g+	g+	t	-179.1	16.8	178.0	17.8	-177.6	13.7	-178.3	11.8	-179.2	18.8
	g+	g+	g+	-68.0	17.4	-70.7	17.2	-72.3	16.6	-71.7	23.1	-63.7	15.6
	g+	g+	g-	81.2	20.9	73.4	9.6	86.0	5.4	56.6	0.0	86.5	24.7
	g+	g-	t	175.5	21.7	175.7	5.6	-171.6	18.2	-	-	172.7	24.2
	g+	g-	g+	-99.3	23.4	-100.0	21.5	-124.2	0.0	-	-	-72.2	0.0
	g+	g-	g-	71.7	18.7	66.2	0.0	78.9	21.8	-	-	65.5	14.0
	g-	t	t	-179.9	19.4	178.7	23.4	-176.7	17.6	177.9	32.3	178.8	17.2
	g-	t	g+	-71.2	21.6	-69.4	25.3	-60.6	9.1	-65.8	4.4	-76.7	21.8
	g-	t	g-	71.7	20.0	69.7	17.8	70.7	21.3	-	-	73.3	19.4
	g-	g+	t	-143.6	0.9	-144.5	0.0	-	-	-	-	-142.7	0.0
	g-	g+	g+	-77.1	3.3	-	-	-	-	-	-	-77.1	3.3
	g-	g+	g-	88.6	0.0	-	-	-	-	-	-	88.6	0.0
	g-	g-	t	-179.3	13.8	165.5	0.0	-168.5	0.0	-	-	-178.2	14.0
Lys	t	t	t	177.8	18.9	177.1	18.7	-179.5	18.2	-180.0	9.5	175.8	20.3
	t	t	g+	-70.9	18.2	-73.1	17.2	-71.6	15.1	-109.2	0.4	-61.8	18.3
	t	t	g-	71.4	17.7	72.5	19.5	69.5	13.7	-	-	70.9	16.5
	t	g+	t	-178.3	24.4	-178.8	26.9	-179.2	6.5	178.4	15.9	-176.6	25.1
	t	g+	g+	-71.4	16.0	-75.5	14.9	-73.6	0.0	-	-	-61.1	16.4
	t	g+	g-	105.3	0.0	105.3	0.0	-	-	-	-	-	-
	t	g-	t	179.7	19.3	179.8	19.8	176.6	11.3	-170.3	20.3	179.7	22.5
	t	g-	g+	-82.4	15.4	-82.0	18.6	-83.0	4.4	-	-	-	-
	t	g-	g-	76.1	16.5	76.6	12.1	94.7	7.4	67.1	21.9	62.9	8.1
	g+	t	t	179.4	19.3	179.8	18.7	-179.4	19.2	177.2	21.6	178.8	19.3
	g+	t	g+	-75.2	21.3	-73.7	20.5	-81.0	24.7	-67.6	21.1	-75.2	20.5
	g+	t	g-	69.8	16.3	70.9	15.4	68.6	14.7	67.1	15.9	69.5	17.7
	g+	g+	t	-177.2	21.2	179.6	19.7	-176.4	19.8	-178.0	19.1	-175.6	22.3
	g+	g+	g+	-73.9	18.5	-75.6	20.5	-70.3	17.8	-87.1	18.5	-72.2	16.0
	g+	g+	g-	88.6	17.1	93.7	16.5	86.1	19.5	-	-	87.1	15.8
	g+	g-	t	177.6	22.0	173.9	20.8	174.6	13.8	179.4	8.1	-178.2	25.5
	g+	g-	g+	-91.2	16.3	-105.4	0.0	-84.0	0.0	-	-	-87.6	19.9
	g+	g-	g-	76.3	17.3	-	-	63.4	0.0	69.7	0.0	79.1	18.7
	g-	t	t	-178.2	19.4	-176.5	9.3	-176.6	19.8	-173.5	9.8	-179.5	21.2
	g-	t	g+	-68.3	20.0	-71.3	21.2	-79.9	19.2	-70.5	0.2	-61.0	18.4
	g-	t	g-	70.4	15.7	62.2	14.6	78.5	12.5	81.3	0.0	77.7	12.9
	g-	g+	t	-178.5	26.6	-	-	-	-	-	-	-178.5	26.6
	g-	g-	t	170.0	30.2	-167.5	13.5	-	-	-	-	147.5	25.1
Met	t	t	t	178.1	23.5	-172.5	20.5	174.1	22.5	-	-	170.4	24.1
	t	t	g+	-71.5	20.6	-67.0	23.1	-73.9	20.7	-89.4	4.1	-69.0	7.8
	t	t	g-	76.8	18.0	73.4	13.3	82.4	19.9	37.6	0.0	74.3	16.4
	t	g+	t	170.4	16.1	-	-	170.4	16.1	-	-	-	-
	t	g+	g+	-76.8	11.4	-74.3	7.2	-78.4	13.0	-	-	-71.7	0.0
	t	g-	t	-172.8	24.3	-173.9	27.4	-177.2	22.5	-133.4	0.0	-172.7	12.0
	t	g-	g+	-89.5	26.1	-84.9	27.9	-112.2	0.5	-	-	-76.5	0.0
	t	g-	g-	74.2	17.0	73.2	14.6	70.4	13.8	63.8	0.0	78.5	21.3

ISBN:978-84-694-2163-5/DL:T.885-2011

Appendix Table 4. *Continue.*

g+	t	t	-176.6	19.7	-175.2	17.2	-176.7	18.7	135.5	0.0	-176.8	22.2
g+	t	g+	-74.4	18.3	-74.1	18.7	-77.1	16.8	-	-	-71.9	19.0
g+	t	g-	74.4	18.3	72.7	17.1	77.2	17.5	71.7	27.0	75.4	19.4
g+	g+	t	163.3	21.8	165.4	21.9	170.5	22.3	143.5	0.0	157.0	19.8
g+	g+	g+	-68.4	15.8	-66.8	16.2	-69.8	17.8	-75.1	7.4	-70.9	12.2
g+	g+	g-	102.7	12.4	104.0	12.7	101.6	0.0	-	-	100.7	12.0
g+	g-	t	157.1	0.0	-	-	157.1	0.0	-	-	-	-
g+	g-	g+	-103.1	0.0	-103.1	0.0	-	-	-	-	-	-
g+	g-	g-	63.5	10.7	-	-	67.5	0.0	-	-	62.1	12.1
g-	t	t	-178.6	18.6	-160.9	0.0	175.9	11.9	179.5	0.0	-174.3	23.9
g-	t	g+	-76.8	19.2	-64.8	23.8	-69.9	15.0	-76.8	0.6	-86.8	19.7
g-	t	g-	75.3	17.6	76.8	16.1	73.8	10.2	-	-	76.1	23.2
g-	g+	t	-169.3	26.0	-143.3	0.0	-	-	-	-	164.8	0.0
g-	g+	g+	-53.1	27.2	-62.9	15.7	-	-	-	-	-43.2	32.1
g-	g-	t	155.0	11.6	-	-	155.0	11.6	-	-	-	-
g-	g-	g+	-116.6	0.0	-	-	-	-	-116.6	0.0	-	-
g-	g-	g-	72.5	6.3	-	-	-	-	-	-	72.5	6.3

The table shows the mean χ_3 values (first column) and their corresponding standard deviations (second column) for the different rotamers in the 1996 "full rotamer libraries" (except for rotamers labelled as "Other"). The data in the table were obtained by using, when possible, combinations of side-chain dihedral angles from χ_1 to χ_3 in nondependent (Averaged) and dependent (Helix, β-Strand, Turn and Coil) secondary-structure coordinates. Residues grouped as "Other" (see **Appendix Table 1**) were not taken into account because they are not a homogeneous sample. • The χ_3 data from Gln and Glu were not used to build the "full rotamer libraries" because their dihedral angle distributions do not have well separated peaks. Consequently, results for Gln and Glu do not appear in this table.

UNIVERSITAT ROVIRA I VIRGILI
ESTUDIS COMPUTACIONALS SOBRE BETA-AMILASA, UNA MOLÈCULA ENZIMÀTICA TIPUS BARRIL (BETA/ALFA) 8: RELACIONS EVOLUTIVES I TRANSICIONS ESTRUCTURALS AL CENTRE CATALÍTIC.
Gerard Pujadas Anguiano
ISBN:978-84-694-2163-5/DL:T.885-2011

Bibliografia.

Aquesta secció de la tesi conté la bibliografia emprada en la elaboració de la introducció, els dos articles ja publicats i l'apèndix.

Adachi, M., Katsube, T., Mikami, B. i Utsumi, S. (1995). Enzymatic and crystallographic analysis of soybean β-amylase expressed in *E. coli*. Publicat directament a DDBJ.

Anfinsen, C.B. i Scheraga, H.A. (1975). Experimental and theoretical aspects of protein folding. *Adv. Protein Chem.* **29**, 205-300.

Attwood, T.K. i Beck, M.E. (1994). PRINTS - A protein motif fingerprint database. *Protein Eng.* **7**, 841-848.

Attwood, T.K., Beck, M.E., Bleasby, A.J. i Parry-Smith, D.J. (1994). PRINTS - A database of protein motif fingerprints. *Nucleic Acids Res.* **22**, 3590-3596.

Bainbridge, G., Madgwick, P., Parmar, S., Mitchell, R., Paul, M., Pitts, J., Keys, A.J. i Parry, M.A.J. (1995). Engineering rubisco to change its catalytic properties. *J. Exp. Bot.* **46**, 1269-1276.

Bairoch, A. (1994). PROSITE: recent developments. *Nucleic Acids Res.* **22**, 3578-3580.

Bairoch, A. i Boeckmann, B. (1994). The Swiss-Prot protein sequence data bank: current status. *Nucleic Acids Res.* **22**, 3578-3580.

Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., Pogson, C.I., Wilson, I.A., Corran, P.H., Furth, A.J., Milman, J.D., Offord, R.E., Priddle, J.D. i Waley, S.G. (1975). Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 Å resolution using amino acid sequence data. *Nature* **255**, 609-614.

Beauregard, M., Goraj, K., Goffin, V., Heremans, K., Goormaghtigh, E., Ruysschaert, J.M. i Martial, J.A. (1991). Spectroscopic investigation of structure in octarellin (a *de novo* protein designed to adopt the α/β-barrel packing). *Protein Eng.* **4**, 745-749.

ISBN:978-84-694-2163-5/DL:T.885-2011

Benedetti, E., Morelli, G., Némethy, G. i Scheraga, H.A. (1983). Statistical and energetic analysis of side-chain conformations in oligopeptides. *Int. J. Pept. Protein Res.* **22**, 1-15.

Benson, D.A., Boguski, M., Lipman, D.J. i Ostell, J. (1994). GenBank. *Nucleic Acids Res.* **22**, 3441-3444.

Bernfeld, P. (1955). Amylases, α and β. *Methods Enzymol.* **1**, 149-158.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. i Tasumi, M. (1977). The Protein Data Bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.

Bhat, T.N., Sasisekharan, V. i Vijayan, M. (1979). An analysis of side-chain conformation in proteins. *Int. J. Pept. Protein Res.* **13**, 170-184.

Birktoft, J.J i Blow, D.M. (1972). Structure of crystalline α-chymotrypsin. *J. Mol. Biol.* **68**, 187-240.

Blaber, M., Zhang, X.J., Lindstrom, J.D., Pepiot, S.D., Baase, W.A. i Matthews, B.W. (1994). Determination of α-helix propensity within the context of a folded protein - sites 44 and 131 in bacteriophage T4 lysozyme. *J. Mol. Biol.* **235**, 600-624.

Bleasby, A.J., Akrigg, D. i Attwood, T.K. (1994). OWL - A non-redundant, composite protein sequence database. *Nucleic Acids Res.* **22**, 3574-3577.

Borchert, T.V., Abagyan, R., Jaenicke, R. i Wierenga, R.K. (1994). Design, creation, and characterization of a stable, monomeric triosephosphate isomerase. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1515-1518.

Borchert, T.V., Abagyan, R., Kishan, K.V.R., Zeelen, J.P. i Wierenga, R.K. (1993). The crystal structure of an engineered monomeric triosephosphate isomerase, monotim - the correct modelling of an eight-residue loop. *Structure* **1**, 205-213.

Bork, P., Gellerich, J., Groth, H., Hooft, R. i Martin F. (1995). Divergent evolution of a β/α-barrel subclass: Detection of numerous phosphate-binding sites by motif search. *Protein Sci.* **4**, 268-274.

Bower, M.J., Cohen, F.E. i Dunbrack, R.L. (1997). Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: A new homology modeling tool. *J. Mol. Biol.* **267**, 1268-1282.

Brändén, C.I. i Tooze, J. (1991). α/β structures. A: "Introduction to Protein Structure". NewYork: Garland Publishing, 43-57.

Bureau, D. i Daussant, J. (1983). Efficiency of a smooth description procedure for the purification of barley β-amylase using immunoaffinity chromatography. *Biochimie* **65**, 361-365.

Castonguay, L.A., Bryant, S.H., Snow, P.M. i Fetrow J.S. (1995). A proposed structural model of domain 1 of fasciclin III neural cell adhesion protein based on an inverse folding algorithm. *Protein Sci.* **4**, 472-483.

Claverie, J.M. (1993). Detecting frameshifts by amino acid sequence comparison. *J. Mol. Biol.* **234**, 1140-1157.

Cleland, W.W. i Kreevoy, M.M. (1994). Low-barrier hydrogen bonds and enzymic catalysis. *Science* **264**, 1887-1890.

Cohen, C. i Parry, D.A.D. (1990). α-helical coiled coils and bundles: how to design and α-helical protein. *Proteins* **7**,1-15.

Coulson, A.F.W. (1994). A proposed structure for 'Family 18' chitinases - A possible function for narbonin. *FEBS Lett.* **354**, 41-44.

Cregut, D., Liautard, J.P. i Chiche, L. (1994). Homology modelling of annexin I: implicit solvation improves side-chain prediction and combination of evaluation criteria allows recognition of different types of conformational error. *Protein Eng.* **7**, 1333-1344.

Creighton, T.E. (1988). Disulphide bonds and protein stability. *Bioessays* **8**, 57-63.

Chan, H.S. i Dill, K.A. (1989). Intrachain loops in polymers: effects of excluded volume. *J. Chem. Phys.* **90**, 492-509.

Chandrasekaran, R. i Ramachandran, G.N. (1970). Studies on the conformation of amino acids. XI. Analysis of the observed side group conformation in proteins. *Int. J. Protein Res.* **2**, 223-233.

Chen, J. (1993). Cloning and characterization of two rice β-amylase genomic clones. Sense publicar.

Cheong, C.G., Eom, S.H., Chang, C., Shin, D.H., Song, H.K., Min, K., Moon, J.H., Kim, K.K., Hwang, K.Y., Suh, S.W. (1995). Crystallization, molecular replacement solution and refinement of tetrameric β-amylase from sweet potato. *Proteins* **21**, 105-117.

Chinea, G., Padron, G., Hooft, R.W.W., Sander, C. i Vriend, G. (1995). The use of position-specific rotamers in model building by homology. *Proteins* **23**, 415-421.

Chothia, C. (1988). The 14th barrel rolls out. *Nature* **333**, 598-599.

Chothia, C. i Finkelstein, A.V. (1990). The classification and origins of protein folding patterns. *Annu. Rev. Biochem.* **59**, 1007-1039.

Chothia, C. i Janin, J. (1982). Orthogonal packing of β-sheets in proteins. *Biochemistry* **21**, 3955-3965.

Dahiyat, B.I. i Mayo, S.L. (1996). Protein design automation. *Protein Sci.* **5**, 895-903.

Darby, N. i Creighton, T.E. (1995). Disulfide bonds in protein folding and stability. A: "Protein Stability and Folding (Series: Methods in Molecular Biology)" **40**, 219-252.

Daussant, J. i Laurière, C. (1990). Detection and partial characterization of two antigenically distinct β-amylases in developing kernels of wheat. *Planta* **181**, 505-511.

Daussant, J., Sadowski, J. i Ziegler, P. (1994). Cereal β-amylases: diversity of the β-amylase isozyme status within cereals. *J. Plant Physiol.* **143**, 585-590.

Daussant, J., Sadowski, J., Rorat, T., Mayer, C. i Laurière, C. (1991). Independent regulatory aspects and post-translational modifications of two β-amylases of rye. Use of a mutant inbred line. *Plant Physiol.* **96**, 84-90.

Davies, G. i Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure* **3**, 853-859.

De Maeyer, M., Desmet, J. i Lasters, I. (1997). All in one: a highly detailed rotamer library improves both accuracy and speed in the modelling of sidechains by dead-end elimination. *Fold Des* **2**, 53-66.

DeGrado, W.F. (1988). Design of peptides and proteins. *Adv. Protein Chem* **39**, 51-124.

DeGrado, W.F.; Wassermann, Z.R. i Lear, J.D. (1989). Protein design, a minimalist approach. *Science* **243**, 622-628.

Desjarlais, J.R. i Handel, T.M. (1995). De novo design of the hydrophobic cores of proteins. *Protein Sci.* **4**, 2006-2018.

Desmet, J., De Maeyer, M. i Lasters, I. (1994). The "Dead-End-Elimination" theorem. A new approach to the side-chain packing problem. A: "The protein folding problem and tertiary structure prediction". Boston: Birkhäuser (Merz, K.Jr. & Le Grand, S., eds), 307-337.

Desmet, J., De Maeyer, M., Hazes, B. i Lasters, I. (1992). The Dead-End elimination theorem and its use in protein side-chain positioning. *Nature* **356**, 539-542.

Doig, A.J. (1996). Thermodynamics of amino acid side-chain internal rotations. *Biophys. Chem.* **61**, 131-141.

Doig, A.J. i Sternberg, M.J.E. (1995). Side-chain conformational entropy in protein folding. *Protein Sci.* **4**, 2247-2251.

Doig, A.J., MacArthur, M.W., Stapley, B.J. i Thornton, J.M. (1997). Structures of N-termini of helices in proteins. *Protein Sci.* **6**, 147-155.

Dominguez, R., Souchon, H., Lascombe, M. i Alzari, P.M. (1996). The crystal structure of a family 5 endoglucanase mutant in complexed and uncomplexed forms reveals an induced fit activation mechanism. *J. Mol. Biol.* **257**, 1042-1051.

Dominguez, R., Souchon, H., Spinelli, S., Dauter, Z., Wilson, K.S., Chauvaux, S., Beguin, P. i Alzari, P.M. (1995). A common protein fold and similar active site in two distinct families of β-glycanases. *Nat. Struct. Biol.* **2**, 569-576.

Dunbrack, R.L. (1993). Conformational analysis of protein sidechains: empirical energy parameters for proline and development of a backbone-dependent rotamer library. PhD thesis, Department of Chemistry, Harvard University, USA.

Dunbrack, R.L. i Cohen, F.E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **6**, 1661-1681.

Dunbrack, R.L. i Karplus, M. (1993). Backbone-dependent rotamer library for proteins - application to side-chain prediction. *J. Mol. Biol.* **230**, 543-574.

Dunbrack, R.L. i Karplus, M. (1994). Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. *Nature Struct. Biol.* **1**, 334-340.

Eder, J. i Wilmanns, M. (1992). Protein engineering of a disulfide bond in a β/α-barrel protein. *Biochemistry* **31**, 4437-4444.

Ehrlich, K.C. i Montalbano, B.G. (1992). Nucleotide and deduced protein sequence of β-amylase. Publicat directament a DDBJ.

Emmert, D.B., Stoehr, P.J., Stoesser, G. i Cameron, G. (1994). The European Bioinformatics Institute (EBI) databases. *Nucleic Acids Res.* **22**, 3445-3449.

Essen, L.,O., Perisic, O., Lynch, D.E., Katan, M. i Williams, R.L. (1997). A ternary metal binding site in the C2 domain of phosphoinositide specific phospholipase cδ1. *Biochemistry* **36**, 2753-2762.

Farber, G.K. (1993). An α/β-barrel full of evolutionary trouble. *Curr. Opin. Struct. Biol.* **3**, 409-412.

Farber, G.K. i Petsko, G.A. (1990). The evolution of α/β barrel enzymes. *Trends Biochem. Sci.* **15**, 228-234.

Franks, F. (1988). Introduction. A: "Characterization of proteins". Cambridge (UK): Pafra Ltd. (Franks, F., ed), 1-8.

Friedberg, F. i Rhodes, C. (1988). Segments of amino acid sequence similarity in β-amylases. *Protein Seq. Data Anal.* **1**, 499-501.

Fukazawa, C. (1994). Method for separation of β-amylase. Patent number 05294341. Tsukuba (Japan).

García-Vallvé, S., Rojas, A., Palau, J. i Romeu, A. (1998). Circular permutants in β-glucosidases (Family 3) within a predicted double-domain topology that includes a $(\beta/\alpha)_8$-barrel. *Proteins*, en premsa.

Garnier, J., Osguthorpe, D.J. i Robson, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* **120**, 97-120.

George, D.G., Barker, W.C., Mewes, H.W., Pfeiffer, F. i Tsugita, A. (1994). The PIR-International protein sequence database. *Nucleic Acids Res.* **22**, 3569-3573.

Gertler, A. i Birk, Y. (1965). Purification and characterization of a β-amylase from soybeans. *Biochem. J.* **95**, 621-627.

Gish, W. i States, D.J. (1993). Identification of protein coding regions by database similarity search. *Nat. Genet.* **3**, 266-272.

Goldstein, R.F. (1994). Efficient rotamer elimination applied to protein side-chains and related spin glasses. *Biophys. J.* **66**, 1335-1340.

Gómez, J., Hilser, V.J., Xie, D. i Freire, E. (1995). The heat capacity of proteins. *Proteins* **22**, 404–412.

Goraj, K., Renard, A. i Martial, J.A. (1990). Synthesis, purification and initial structural characterization of octarellin, a *de novo* polypeptide modelled on the α/β-barrel proteins. *Protein Eng.* **3**, 259-266.

Gray, T.M. i Matthews, B.W. (1984). Intrahelical hydrogen bonding of serine, threonine and cysteine residues within α-helices and its relevance to membrane-bound proteins. *J. Mol. Biol.* **175**, 75-81.

Gregoret, L.M., Rader, S.D., Fletterick, R.J. i Cohen, F.E. (1991). Hydrogen bonds involving sulfur atoms in proteins. *Proteins* **9**, 99-107.

Gunasekaran, K., Ramakrishnan, C. i Balaram, P. (1996). Disallowed Ramachandran conformations of amino acid residues in protein structures. *J. Mol. Biol.* **264**, 191-198.

Harbury, P.B., Tidor, B. i Kim, P.S. (1995). Repacking protein cores with backbone freedom: structure prediction for coiled coils. *Proc. Natl Acad. Sci. USA,* **92**, 8408-8412.

ISBN:978-84-694-2163-5/DL:T.885-2011

Hecht, M.H., Richardson, J.S., Richardson, D.C. i Ogden, R.C. (1990). *De novo* design, expression, and characterization of Felix: a four-helix bundle protein of native-like sequence. *Science* **249**, 884-891.

Helin, S., Kahn, P.C., Guha, B.L., Mallows, D.G. i Goldman, A. (1995). The refined X-ray structure of muconate lactonizing enzyme from *Pseudomonas putida* PRS2000 at 1.85 Å resolution. *J. Mol. Biol.* **254**, 918-941

Hennig, M., Jansonius, J.N., Terwisscha van Scheltinga, A.C., Dijkstra, B.W. i Schlesier, B. (1995a). Crystal structure of concanavalin B at 1.65 Å resolution. An "inactivated" chitinase from seeds of *Canavalia ensiformis*. *J. Mol. Biol.* **254**, 237-246.

Hennig, M., Pfefferhennig, S., Dauter, Z., Wilson, K.S., Schlesier, B. i Nong, V.H. (1995b). Crystal structure of narbonin at 1.8 Å resolution. *Acta Cryst.* **B51**, 177-189.

Hennig, M., Schlesier, B., Dauter, Z., Pfeffer, S., Betzel, C., Hohne, W.E. i Wilson, K.S. (1992). A TIM barrel protein without enzymatic activity? - Crystal-structure of narbonin at 1.8 Å resolution. *FEBS Lett.* **306**, 397-408.

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino-acid sequence similarities. *Biochem. J.* **280**, 309-316.

Henrissat, B. i Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **293**, 781-788.

Henrissat, B. i Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.* **316**, 695-696.

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.P. i Davies, G. (1995). Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7090-7094.

Herzberg, O. i Moult, J. (1991). Analysis of the steric strain in the polypeptide backbone of protein molecules. *Proteins* **11**, 223-229.

Higgins, D.G. i Sharp, P.M. (1989). Fast and sensitive multiple sequence alignments on a microcomputer. *Comput. Appl. Biosci.* **5**, 151-153.

Hobohm, U. i Sander, C. (1994). Enlarged representative set of protein structures. *Protein Sci.* **3**, 522-524.

Hobohm, U., Scharf, M., Schneider, R. i Sander, C. (1992). Selection of a representative set of structures from the Brookhaven Protein Data Bank. *Protein Sci.* **1**, 409-417.

Hooft, R.W.W., Sander, C., Scharf, M. i Vriend, G. (1996). The PDBFINDER database: a summary of PDB, DSSP and HSSP information with added value. *CABIOS* **12**, 525-529.

Houbrechts, A., Moreau, B., Abagyan, R., Mainfroid, V., Preaux, G., Lamproye, A., Poncin, A., Goormaghtigh, E., Ruysschaert, J.M., Martial, J.A. i Goraj, K. (1995). Second-generation octarellins: two new *de novo* $(\beta/\alpha)_8$ polypeptides designed for investigating the influence of β-residue packing on the α/β-barrel structure stability. *Protein Eng.* **8**, 249-259.

Hsieh, M., Hensley, P., Brenowitz, M. i Fetrow, J.S. (1994). A molecular model of the inducer binding domain of the galactose repressor of *Escherichia coli*. *J. Biol. Chem.* **269**, 13825-13835.

Hwang, K.Y., Song, H.K., Chang, C., Lee, J., Lee, S.Y., Kim, K.K., Choe, S., Sweet, R.M. i Suh, S.W. (1997). Crystal structure of thermostable α amylase from *Bacillus licheniformis* refined at 1.7 Å resolution. *Mol. Cells.* **7**, 251-258.

IUPAC-IUB Commission on Biochemical nomenclature. (1970). Abbreviations and symbols for the description of the conformation of polypeptide chains. Tentative rules (1969). *Biochemistry* **9**, 3471-3479.

Iyer, L.K. i Vishveshwara, S. (1996). The stability of transmembrane helices: a molecular dynamics study on the isolated helices of bacteriorhodopsin. *Biopolymers* **38**, 401-421.

James, M.N. i Sielecki, A.R. (1983). Structure and refinement of penicillopepsin at 1.8 Å resolution. *J. Mol. Biol.* **163**, 299-361.

Janecek, S. (1993). Does the increased hydrophobicity of the interior and hydrophilicity of the exterior of an enzyme structure reflect its increased thermostability?. *Int. J. Biol. Macromol.* **15**, 317-318

Janecek, S. (1994). Parallel β/α-barrels of α-amylase, cyclodextrin glycosyltransferase and oligo-1,6-glucosidase versus the barrel of β-amylase:

evolutionary distance is a reflection of unrelated sequences. *FEBS Lett.* **353**, 119-123.

Janecek, S. (1995). Similarity of different β-strands flanked in loops by glycines and prolines from distinct $(\alpha/\beta)_8$-barrel enzymes: Chance or a homology?. *Protein Sci.* **4**, 1239-1242.

Janecek, S. (1996). Invariant glycines and prolines flanking in loops the strand β2 of various $(\alpha/\beta)_8$-barrel enzymes: A hidden homology?. *Protein Sci.* **5**, 1136-1143.

Janecek, S. i Balaz, S. (1995). Functionally essential, invariant glutamate near the C-terminus of strand β5 in various $(\alpha/\beta)_8$-barrel enzymes as a possible indicator of their evolutionary relatedness. *Protein Eng.* **8**, 809-813.

Janecek, S. i Bateman, A. (1996). The parallel $(\alpha/\beta)_8$-barrel: perhaps the most universal and the most puzzling protein folding motif. *Biologia (Bratisl)* **51**, 613–628.

Janin, J., Wodak, S., Levitt, M. i Maigret, B. (1978). Conformation of amino acid side-chains in proteins. *J. Mol. Biol.* **125**, 357-386.

Ji, E.S., Mikami, B., Kim, J.P. i Morita, Y. (1990). Positions of substituted amino acids in soybean β-amylase isoenzymes. *Agric. Biol. Chem.* **54**, 3065-3067.

Jia, J., Huang, W.J., Schorken, U., Sahm, H., Sprenger, G.A., Lindqvist, Y. i Schneider, G. (1996). Crystal structure of transaldolase B from *Escherichia coli* suggests a circular permutation of the α/β barrel within the class I aldolase family. *Structure* **4**, 715-724

Kabsch, W. i Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers,* **22**, 2577-2637.

Kanaya, S., Katsuda, C., Kimura, S., Nakai, T., Kitakuni, E., Nakamura, H., Katayanagi, K., Morikawa, K. i Ikehara, M. (1991). Stabilization of *Escherichia coli* ribonuclease H by introduction of an artificial disulfide bond. *J. Biol. Chem.* **226**, 6038-6044.

ISBN:978-84-694-2163-5/DL:T.885-2011

Karlin, S., Bucher, P., Brendel, V. i Altschul, S.F. (1991). Statistical methods and insights for protein and DNA sequences. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 175-203.

Kawazu, T., Nakanishi, Y., Uozumi, N., Sasaki, T., Yamagata, H., Tsukagoshi, N. i Udaka, S. (1987). Cloning and nucleotide sequence of the gene coding for enzymatically active fragments of the *Bacillus polymyxa* β-amylase. *J. Bacteriol.* **169**, 1564-1570.

Ke, H. (1992). Similarities and differences between human cyclophilin A and other β-barrel structures. *J. Mol. Biol.* **228**, 539-550.

Keller, D.A., Shibata, M., Marcus, E., Ornstein, R.L. i Rein, R. (1995). Finding the global minimum: A fuzzy end elimination implementation. *Protein Eng.* **8**, 893-904.

Kimura, M. (1983). "The neutral theory of molecular evolution". Cambridge: Cambridge University Press.

Kishan, R., Zeelen, J.P., Noble, M.E.M., Borchert, T.V., Mainfroid, V., Goraj, K., Martial, J.A. i Wierenga, R.K. (1994). Modular mutagenesis of a TIM-barrel enzyme: the crystal structure of a chimeric *E. coli* TIM having the eighth βα-unit replaced by the equivalent unit of chicken TIM. *Protein Eng.* **7**, 945-951.

Kitamoto, N., Yamagata, H., Kato, T., Tsukagoshi, N. i Udaka, S. (1988). Cloning and sequencing of the gene encoding thermophilic β-amylase of *Clostridium thermosulfurogenes*. *J. Bacteriol.* **170**, 5848-5854.

Koehl, P. i Delarue, M. (1994). Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. *J. Mol. Biol.* **239**, 249-275.

Koehl, P. i Delarue, M. (1995). A self consistent mean field approach to simultaneous gap closure and side-chain positioning in homology modelling. *Nat. Struct. Biol.* **2**, 163-170.

Kohno, A., Shinke, R. i Nanmori, T. (1990). Features of the β-amylase isoform system in dry and germinating seeds of alfalfa (*Medicago sativa* L). *Biochim. Biophys. Acta* **1035**, 325-330.

Kono, H. i Doi, J. (1994). Energy minimization method using automata network for sequence and side-chain conformation prediction from given backbone geometry. *Proteins* **19**, 244-255.

Kotelchuck, D. i Scheraga, H.A. (1968). The influence of short-range interactions on protein conformation. I. Side chain-backbone interactions within a single peptide unit. *Proc. Natl. Acad. Sci. U.S.A.*, **61**, 1163-1170.

Kreis, M., Williamson, M., Buxton, B., Pywell, J., Hejgaard, J. i Svendsen, I. (1987). Primary structure and differential expression of β-amylase in normal and mutant barleys. *Eur. J. Biochem.* **169**, 517-525.

Kuma, K., Iwabe, N. i Miyata, T. (1995). Functional constraints against variations on molecules from the tissue level: slowly evolving brain-specific genes demonstrated by protein kinase and immunoglobulin supergene families. *Mol. Biol. Evol.* **12**, 123-130.

Kuntz, I.D., Meng, E.C. i Schoichet, B.K. (1994). Structure-based molecular design. *Acc. Chem. Res.* **27**, 117–123.

Kyte, J. i Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105-132.

LaBerge, D.E. i Meredith, O.S. (1969). Chromatographic properties of barley and malt β-amylases. *J. Inst. Brew.* **75**, 19-25.

Lake, J.A. (1994). Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1455-1459.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. i Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291.

Lasters, I. (1990). Estimating the twist of β-strands embedded within a regular parallel β-barrel structure. *Protein Eng.* **4**, 133-135.

Lasters, I., De Maeyer, M. i Desmet, J. (1995). Enhanced dead-end elimination in the search for the global minimum energy conformation of a collection of protein side chains. *Protein Eng.* **8**, 815-822.

ISBN:978-84-694-2163-5/DL:T.885-2011

Lasters, I., Wodak, S.J., Alard, Ph. i van Cutsem, E. (1988). Structural principles of parallel β-barrels in proteins. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3338-3342.

Lasters, I., Wodak, S.J., Pio, F. (1990). The design of idealized α/β barrels: analysis of β-sheet closure requirements. *Proteins* **7**, 249-256.

Leach, A.R. (1994). Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.* **235**, 345-356.

Leach, S.J., Nemethy, G. i Scheraga, H.A. (1966). Computation of the sterically allowed conformations of peptides. *Biopolymers,* **4**, 369-407.

Lebioda, L., Stec, B. i Brewer J.M. (1989). The structure of yeast enolase at 2.25 Å resolution. An 8-fold β+α barrel with a novel ββαα$(\beta\alpha)_6$ topology. *J. Biol. Chem.* **264**, 3685-3693.

Lesk, A.M., Brändén, C-I. i Chothia, C. (1989). Structural principles of α/β barrel proteins: the packing of the interior of the sheet. *Proteins* **5**, 139-148.

Lin, T.P., Spilatro, S.R. i Preiss, J. (1988). Subcellular location and characterization of amylases in *Arabidopsis* leaf. *Plant Physiol.* **86**, 251-259.

Lindqvist, Y., Brändén, C.I., Mathews, F.S. i Lederer, F. (1991). Spinach glycolate oxidase and yeast flavocytochrome b2 are structurally homologous and evolutionarily related enzymes with distinctly different function and flavin mononucleotide binding. *J. Biol. Chem.* **266**, 3198-3207.

Lipkind, G.M. i Popov, E.M. (1971). Conformational states of amino acid residues in proteins. Side chains. *Mol. Biol.* **5**, 532-542.

Lonberg, N. i Gilbert, W. (1985). Intron/exon structure of the chicken pyruvate kinase gene. *Cell* **40**, 81-90.

Luger, K., Hommel, U., Herold, M., Hofsteenge, J. i Kirschner, K. (1989). Correct folding of circularly permuted variants of a βα barrel enzyme in vivo. *Science* **243**, 206-210.

Luger, K., Szadkowski, H. i Kirschner, K. (1990). An 8-fold βα barrel protein with redundant folding possibilities. *Protein Eng.* **3**, 249-258.

Lundgard, R. i Svensson, B. (1987). The four major forms of barley β-amylase. Purification, characterization and structural relationship. *Carlsberg Res. Commun.* **52**, 313-326.

MacGregor, E.A., Jespersen, H.M. i Svensson, B. (1996). A circularly permuted α-amylase-type α/β-barrel structure in glucan-synthesizing glucosyltransferases. *FEBS Lett.* **378**, 263-266.

Mainfroid, V., Goraj, K., Rentierdelrue, F., Houbrechts, A., Loiseau, A., Gohimont, A.C., Noble, M.E.M., Borchert, T.V., Wierenga, R.K. i Martial, J.A. (1993). Replacing the (βα)-unit 8 of *E. coli* TIM with its chicken homologue leads to a stable and active hybrid enzyme. *Protein Eng.* **6**, 893-900.

Mainfroid, V., Mande, S.C., Hol, W.G.J., Martial, J.A. i Goraj, K. (1996). Stabilization of human triosephosphate isomerase by improvement of the stability of individual α-helices in dimeric as well as monomeric forms of the protein. *Biochemistry* **35**, 4110-4117.

Mark, A.E. i van Gunsteren, W.F. (1994). Decomposition of a free energy of a system in terms of specific interactions. Implications for theoretical and experimental studies. *J. Mol. Biol.* **240**, 167–176.

Matsumura, M., Becktel, W.J., Levitt, M. i Matthews, B.W. (1989b). Stabilization of phage T4 lysozyme by engineered disulfide bonds. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6562-6566.

Matsumura, M., Signor, G. i Matthews, B.W. (1989a). Substantial increase of protein stability by multiple disulphide bonds. *Nature* **342**, 291-293.

McDonald, I.K. i Thornton, J.M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777-793.

McGregor, M.J., Islam, S.A. i Sternberg, M.J.E. (1987). Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *J. Mol. Biol.* **198**, 295-310.

McLachlan, A.D. (1979). Gene duplications in the structural evolution of chymotrypsin. *J. Mol. Biol.* **128**, 49-79.

Meiering, E.M., Serrano, L. i Fersht, A.R. (1992). Effect of active site residues in barnase on activity and stability. *J. Mol. Biol.* **225**, 585-589.

ISBN:978-84-694-2163-5/DL:T.885-2011

Mikami, B. i Morita, Y. (1988). Amino acid sequence and three dimensional structure of β-amylase. A: "Handbook of Amylases and Related Enzymes". Oxford: Pergamon Press (The Amylase Research Society of Japan, ed), 99-104.

Mikami, B. i Morita, Y. (1988). Soybean β-amylase. A: "Handbook of Amylases and Related Enzymes". Oxford: Pergamon Press (The Amylase Research Society of Japan, ed), 87-89.

Mikami, B., Aibara, S. i Morita, Y. (1982). Distribution and propierties of soybean β-amylase isozymes. *Agric. Biol. Chem.* **46**, 943-953.

Mikami, B., Degano M., Hehre E.J. i Sacchettini J.C. (1994). Crystal structures of soybean β-amylase reacted with β-maltose and maltal: active site components and their apparent roles in catalysis. *Biochemistry* **33**, 7779-7787.

Mikami, B., Hehre, E.J., Sato, M., Katsube, Y., Hirose, M., Morita, Y. i Sacchettini, J.C. (1993). The 2.0 Å resolution structure of soybean β-amylase complexed with α-cyclodextrin. *Biochemistry* **32**, 6836-6845.

Mikami, B., Nomura, K. i Morita, Y. (1986). N-terminal sequence of soybean β-amylase. *J. Biochem.* **100**, 513-516.

Mikami, B., Nomura, K. i Morita, Y. (1994). Two sulfhydryl groups near the active site of soybean β-amylase. *Biosci. Biotechnol. Biochem.* **58**, 126-132.

Mikami, B., Sato, M., Shibata, T., Hirose, M., Aibara, S., Katsube, Y. i Morita, Y. (1992). Three-dimensional structure of soybean β-amylase determined at 3.0 Å resolution: preliminary chain tracing of the complex with α-cyclodextrin. *J. Biochem.* **112**, 541-546.

Mitchinson, C. i Wells, J.A. (1989). Protein engineering of disulfide bonds in subtilisin BPN'. *Biochemistry* **28**, 4807-4815.

Monroe, J.D., Salminen, M.D. i Preiss, J. (1991). Nucleotide sequence of a cDNA clone encoding a β-Amylase from *Arabidopsis thaliana*. *Plant Physiol.* **97**, 1599-1601.

Muirhead, H. (1983). Triose phosphate isomerase, pyruvate kinase and other α/β-barrel enzymes. *Trends Biochem. Sci.* **8**, 326-330.

Murzin, A.G., Brenner, S.E., Hubbard, T. i Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536-540.

Murzin, A.G., Lesk, A.M. i Chothia, C. (1992). β-trefoil fold patterns of structure and sequence in the kunitz inhibitors interleukins-1β and 1α and fibroblast growth factors. *J. Mol. Biol.* **223**, 531-543.

Murzin, A.G., Lesk, A.M. i Chothia, C. (1994a). Principles determining the structure of β-sheet barrels in proteins. I. A theoretical analysis. *J. Mol. Biol.* **236**, 1369-1381.

Murzin, A.G., Lesk, A.M. i Chothia, C. (1994b). Principles determining the structure of β-sheet barrels in proteins. II. The observed structures. *J. Mol. Biol.* **236**, 1382-1400.

Nakamura, H., Tanimura, R. i Kidera, A. (1996). Side-chain conformations cooperatively restricted in protein secondary structure. 1. A novel method for exhaustive structure searching. *Proc. Japan Acad.* **72B**, 143-148.

Nanmori, T. (1988). Bacterial β-amylases (*B. cereus*, *B. polymyxa*, etc.). A: "Handbook of Amylases and Related Enzymes". Oxford: Pergamon Press (The Amylase Research Society of Japan, ed), 94-99.

Nanmori, T., Nagai, M., Shimizu, Y., Shinke, R. i Mikami, B. (1993). Cloning of the β-amylase gene from *Bacillus cereus* and characteristics of the primary structure of the enzyme. *Appl. Environ. Microbiol.* **59**, 623-627.

Nayeem, A. i Scheraga, H.A. (1994). A statistical analysis of side-chain conformations in proteins: comparison with ECEPP predictions. *J. Protein Chem.* **13**, 283-296.

NC-IUBMB. (1992). "Enzyme Nomenclature". Academic Press, New-York.

Nitta, Y., Isoda, Y., Toda, H. i Sakiyama, F. (1989). Identification of glutamic acid 186 affinity-labeled by 2,3-epoxypropyl α-D-glucopyranoside in soybean β-amylase. *J. Biochem.* **105**, 573-576

Nomura, K., Yoneda, I., Nanmori, T., Shinke, R., Morita, Y. i Mikami, B. (1995). The role of SH and S-S groups in *Bacillus cereus* β-amylase. *J. Biochem.* **118**, 1124-1130.

Novotny, J., Bruccoleri, R.E. i Newell, J. (1984). Twisted hyperboloid (strophoid) as a model of β-barrels in proteins. *J. Mol. Biol.* **177**, 567-573.

Ogata, K. i Umeyama, H. (1997). Prediction of protein side-chain conformations by principal component analysis for fixed main-chain atoms. *Protein Eng.* **10**, 353-359.

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B. i Thornton, J.M. (1997). CATH - a hierarchic classification of protein domain structures. *Structure* **5**, 1093-1108.

Pattabiraman, N., Namboodiri, K., Lowrey, A. i Gaber, B.P. (1990). NRL-3D: a sequence-structure database derived from the protein data bank (PDB) and searchable within the PIR environment. *Protein Seq. Data Anal.* **3**, 387-405.

Perry, L.J. i Wetzel, R. (1984). Disulfide bond engineered into T4 lysozyme: stabilization of the protein toward thermal inactivation. *Science* **226**, 555-557.

Pickett, S.D. i Sternberg, M.J.E. (1993). Empirical scale of side-chain conformational entropy in protein folding. *J. Mol. Biol.* **231**, 825-839.

Piela, L., Nemethy, G. i Scheraga, H.A. (1987). Conformational constraints of amino acid side chains in α-helices. *Biopolymers,* **26**, 1273-1286.

Ponder, J.W. i Richards, F.M. (1987). Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* **193**, 775-791.

Ponnuswamy, P.K. i Sasisekharan, V. (1970a). Studies on the conformation of amino acids. VII. Backbone and side-chain conformations of N-terminal residues in peptides. *Biochim-Biophys-Acta.* **221**, 153-158.

Ponnuswamy, P.K. i Sasisekharan, V. (1970b). Studies on the conformation of amino acids. VIII. Backbone and side-chain conformations of C-terminal residues in peptides. *Biochim-Biophys-Acta.* **221**, 159-164.

Pujadas, G. i Palau, J. (1997). Anatomy of a conformational transition of β-strand 6 in soybean β-amylase caused by substrate (or inhibitor) binding to the catalytical site. *Protein Sci.* **6**, 2409-2417.

ISBN:978-84-694-2163-5/DL:T.885-2011

Pujadas, G., Ramirez, F.M., Valero, R. i Palau, J. (1996). Evolution of β-amylase: patterns of variation and conservation in subfamily sequences in relation to parsimony mechanisms. *Proteins* **25**, 456-472.

Raina, S. i Missiakas, D. (1997). Making and breaking disulfide bonds. *Annu. Rev. Microbiol.* **51**, 179-202.

Raine, A.R.C., Scrutton, N.S. i Mathews, F.S. (1994). On the evolution of alternate core packing in eightfold β/α-barrels. *Protein Sci.* **3**, 1889-1892.

Ramachandran, G.N. i Lakshminarayanan, A.V. (1966). Conformation of side groups in amino acids and peptides. *Biopolymers,* **4**, 495-497.

Ramachandran, G.N., Mazumdar, S.K., Venkatesan, K. i Lakshminarayanan, A.V. (1966). Conformation of the arginine side-group and its variations. *J. Mol. Biol.* **15**, 232-242.

Raychaudhuri, S., Younas, F., Karplus, P.A., Faerman, C.H. i Ripoll, D.R. (1997). Backbone makes a significant contribution to the electrostatics of α/β-barrel proteins. *Protein Sci.* **6**, 1849-1857.

Reardon, D. i Farber, G.K. (1995). The structure and evolution of α/β barrel proteins. *FASEB J.* **9**, 497-503.

Rhodes, C., Strasser, J. i Friedberg, F. (1987). Sequence of an active fragment of *B. polymyxa* β amylase. *Nucleic Acids Res.* **15**, 3934-3934.

Richardson, D.C. i Richardson, J.S. (1992). The kinemage: a tool for scientific communication. *Protein Sci.* **1**, 3-9.

Richardson, J.S. (1981). The anatomy and taxonomy of protein structures. *Advan. Protein Chem.* **34**, 167-339.

Richardson, J.S. i Richardson, D.C. (1989). The *de novo* design of protein structures. *Trends Biochem. Sci.* **14**, 304-309.

Richardson, J.S., Kneller, D., Osguthorpe, D. i Sharf, M. (1987). Babarellin, a design for a four-fold symmetric TIM barrel. A: "Protein Design Exercises, EMBL Biocomputing Technical (Document 1)". Heidelberg: European Molecular Biology Laboratory (Sander, C., ed), 21-39.

ISBN:978-84-694-2163-5/DL:T.885-2011

Robyt, J.F. i Whelon, W.J. (1968). The β-amylase. A: "Starch and its derivatives". London: Chapman and Hall (Radley, J., ed), 430-476.

Rojas, A. i Romeu, A. (1996). A sequence analysis of the β-glucosidase sub-family B. *FEBS Lett.* **378**, 93-97.

Rorat, T., Sadowski, J., Grellet, F., Daussant, J. i Delseny, M. (1991). Characterization of cDNA clones for rye endosperm β-amylase and analysis of β-amylase deficiency in rye mutant lines. *Theor. Appl. Genet.* **83**, 257-263.

Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J.K.C. i Jones, T.A. (1990). Three–dimensional structure of cellobiohydrolase II from *Trichoderma reesei. Science* **249**, 380-385.

Sadowski, J., Rorat, T., Cooke, R. i Delsensy, M. (1993). Nucleotide sequence of a cDNA clone encoding ubiquitous β-amylase in rye (*Secale cereale* L). *Plant Physiol.* **102**, 315-316.

Sadowski, J., Wiatroszak, I. i Daussant, J. (1986). Variability of β-amylase isozymes within a collection of inbred lines or rye (*Secale cereale* L). *Experientia* **44**, 352-354.

Salemme, F.R. (1981). Conformational and geometrical properties of β-sheets in proteins. III. Isotropically stressed configurations. *J. Mol. Biol.* **146**, 143-156.

Salemme, F.R. i Weatherford, D.W. (1981a). Conformational and geometrical properties of β-sheets in proteins. I. Parallel β-sheets. *J. Mol. Biol.* **146**, 101-117.

Salemme, F.R. i Weatherford, D.W. (1981b). Conformational and geometrical properties of β-sheets in proteins. II. Antiparallel and mixed β-sheets. *J. Mol. Biol.* **146**, 119-141.

Sander, C. (1991). De novo design of proteins. *Curr. Opin. Struct. Biol.* **1**, 630-637.

Sander, C., Vriend, G., Bazan, F., Horovitz, A., Nakamura, H., Ribas, L., Finkelstein, A.V., Lockhart, A., Merkl, R., Perry, L.J., Emery, S.C., Gaboriaud, C., Marks, C., Moult, J., Verlinde, C., Eberhard, M., Elofsson, A., Hubbard, T.J.P., Regan, L., Banks, J., Jappelli, R., Lesk, A.M. i Tramontano, A. (1992). Protein design on computers - five new proteins - Shpilka, Grendel, Fingerclasp, Leather, and Aida. *Proteins* **12**, 105-110.

Sasisekharan, V., Ponnuswamy, P.K. (1970). Backbone and side-chain conformations of amino acids and amino acid residues in peptides. *Biopolymers,* **9**, 1249-1256.

Sasisekharan, V., Ponnuswamy, P.K. (1971). Studies on the conformation of amino acids. X. Conformations of norvalyl, leucyl and aromatic side groups in a dipeptide unit. *Biopolymers,* **10**, 583-592.

Sauer, R.T., Hehir, K., Stearman, R.S., Weiss, M.A., Jeitler-Nilsson, A., Suchanek, E.G. i Pabo, C.O. (1986). An engineered intersubunit disulfide enhances the stability and DNA binding of the N-terminal domain of lambda repressor. *Biochemistry* **25**, 5992-5998.

Sayle, R. i Milner-White, E.J. (1995). RasMol: biomolecular graphics for all. *Trends Biochem. Sci.* **20**, 333-379.

Schrauber, H., Eisenhaber, F. i Argos, P. (1993). Rotamers - to be or not to be - an analysis of amino acid side-chain conformations in globular proteins. *J. Mol. Biol.* **230**, 592-612.

Selvaggini, C., Salmona, M. i De Gioia, L. (1995). Manganese peroxidase from *Phanerochaete chrysosporium.* A homology-based molecular model. *Eur. J. Biochem.* **228**, 955-961.

Shenkin, P.S., Farid, H. i Fetrow, J.S. (1996). Prediction and evaluation of side-chain conformations for protein backbone structures. *Proteins* **26**, 323-352.

Shewry, P.R., Parmar, S., Buxton, B., Gale, M.D., Liu, C.J., Hejgaard, J. i Kreis, M. (1988). Multiple molecular forms of β-amylase in seeds and vegetative tissues of barley. *Planta* **176**, 127-134.

Shoichet, B.K., Baase, W.A., Kuroki, R. i Matthews, B.W. (1995). A relationship between protein stability and protein function. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 452-456.

Siggens, K.W. (1987). Molecular cloning and characterization of the β-amylase gene from *Bacillus circulans. Mol. Microbiol.* **1**, 86-91.

Skinner, R., Abrahams, J.P., Whisstock, J.C., Lesk, A.M., Carrell, R.W. i Wardell, M.R. (1997). The 2.6 Å structure of antithrombin indicates a conformational change at the heparin binding site. *J. Mol. Biol.* **266**, 601-609.

Smith, T.J. (1995). MolView: a program for analyzing and displaying atomic structures on the Macintosh personal computer. *J. Mol. Graphics* **13**, 122-125.

Sternberg, M.J.E. i Chickos, J.S. (1994). Protein side-chain conformational entropy derived from fusion data - comparison with other empirical scales. *Protein Eng.* **7**, 149-155.

Takekawa, S., Uozumi, N., Tsukagoshi, N. i Udaka, S. (1991). Proteases involved in generation of β- and α-amylases from a large amylase precursor in *Bacillus polymyxa*. *J. Bacteriol.* **173**, 6820-6825.

Tanaka, T., Anaguchi, H., Hayashi, M., Fukuhara, K., Hubbard, T.J.P., Nakamura, H. i Ikehara, M. (1991). *De novo* design and the synthesis of TIM barrel proteins. *Proc. 12th Am. Pept. Symp.* 376-377.

Tanaka, T., Kimura, H., Hayashi, M., Fujiyoshi, Y., Fukuhara, K-I. i Nakamura, H. (1994a). Characteristics of a *de novo* designed protein. *Protein Sci.* **3**, 419-427.

Tanaka, T., Hayashi, M., Kimura, H., Oobatake, M. i Nakamura, H. (1994b). *De novo* design and creation of a stable artificial protein. *Biophys. Chem.* **50**, 47-61.

Tanaka, T., Kuroda, Y., Kimura, H., Kidokoro, S. i Nakamura, H. (1994c). Cooperative deformation of a *de novo* designed protein. *Protein Eng.* **7**, 969-976.

Tanimura, R., Kidera, A. i Nakamura, H. (1994). Determinants of protein side-chain packing. *Protein Sci.* **3**, 2358-2365.

Terwisscha van Scheltinga, A.C., Armand, S., Kalk, K.H., Isogai, A., Henrissat, B. i Dijkstra, B.W. (1995). Stereochemistry of chitin hydrolysis by a plant chitinase/lysozyme and X-ray structure of a complex with allosamidin: evidence for substrate assisted catalysis. *Biochemistry* **34**, 15619-15623.

Terwisscha van Scheltinga, A.C., Hennig, M. i Dijkstra, B.W. (1996). The 1.8 Å resolution structure of hevamine, a plant chitinase/lysozyme, and analysis of the conserved sequence and structure motifs of glycosyl hydrolase family 18. *J. Mol. Biol.* **262**, 243-257.

Thoma, J.A., Spradlin, J.E. i Dygert, S. (1971). Plant and animal amylases. A: "The Enzymes". New York: Academic Press (Boyer, P.D., ed), 3rd ed., chap 5, 115-189.

Tkachuk, R. i Tipples, K.H. (1966). Wheat β-amylases. II. Characterization. *Cereal Chem.* **43**, 62-79.

Toda, H., Nitta, Y., Asanami, S., Kim, J.P. i Sakiyama, F. (1993). Sweet potato β-amylase. Primary structure and identification of the active-site glutamyl residue. *Eur. J. Biochem.* **216**, 25-38.

Totsuka, A. i Fukazawa, C. (1996). Functional analysis of Glu_{380} and Leu_{383} of soybean β-amylase. A proposed action mechanism. *Eur. J. Biochem.* **240**, 655-659.

Totsuka, A., Fukazawa, C. (1993). Expression and mutation of soybean β-amylase in *Escherichia coli*. *Eur. J. Biochem.* **214**, 787-794.

Totsuka, A., Nong, V.H., Kadokawa, H., Kim, C.S., Itoh, Y. i Fukazawa, C. 1994. Residues essential for catalytic activity of soybean β-amylase. *Eur. J. Biochem.* **221**, 649-654.

Tsuji, T., Chrunyk, B.A., Chen, X.W. i Matthews, C.R. (1993). Mutagenic analysis of the interior packing of an α/β barrel protein - effects on the stabilities and rates of interconversion of the native and partially folded forms of the α-subunit of tryptophan synthase. *Biochemistry* **32**, 5566-5575.

Tuffery, P., Etchebest, C. i Hazout, S. (1997). Prediction of protein side chain conformations: a study on the influence of backbone accuracy on conformation stability in the rotamer space. *Protein Eng.* **10**, 361-372.

Tuffery, P., Etchebest, C., Hazout, S. i Lavery, R. (1991). A new approach to the rapid determination of protein side chain conformations. *J. Biomol. Struct. Dynam.* **8**, 1267-1289.

Tuffery, P., Etchebest, C., Popot, J.L. i Lavery, R. (1994). Prediction of the positioning of the seven transmembrane α-helices of bacteriorhodopsin - a molecular simulation study. *J. Mol. Biol.* **236**, 1105-1122.

Uozumi, N. (1992). Functional roles of active site residues of *Bacillus polymyxa* β-amylase. *Ann. N.Y. Acad. Sci.* **672**, 24-28.

Uozumi, N., Matsuda, T., Tsukagoshi, N., Udaka, S. (1991). Structural and functional roles of cysteine residues in *Bacillus polymyxa* β-amylase. *Biochemistry* **89**, 4594-4599.

Uozumi, N., Sakurai, K., Sasaki, T., Takekawa, S., Yamagata, H., Tsukagoshi, N. i Udaka, S. (1989). A single gene directs synthesis of a precursor protein with β- and α-amylase activities in *Bacillus polymyxa*. *J. Bacteriol.* **171**, 375-382.

Urfer, R. i Kirschner, K. (1992). The importance of surface loops for stabilizing an eightfold βα barrel protein. *Protein Sci.* **1**, 31-45.

van Gelder, C.W.G., Leusen, F.J.J., Leunissen, J.A.M. i Noordik, J.H. (1994). A molecular dynamics approach for the generation of complete protein structures from limited coordinate data. *Proteins* **18**, 174-185.

van Gunsteren, W.F. i Berendsen, H.J.C. (1987). "BIOMOS B.V. Groningen Molecular Simulation (GROMOS) Library Manual". Groningen (the Netherlands, eds, Nijenborgh 16).

Vásquez, M. (1995). An evaluation of discrete and continuum search techniques for conformational analysis of side chains in proteins. *Biopolymers* **36**, 53-70.

Vásquez, M. (1996). Modeling side-chain conformation. *Curr. Opin. Struct. Biol.* **6**, 217-221.

Viguera, A.R. i Serrano, L. (1995). Side-chain interactions between sulfur-containing amino acids and phenylalanine in α-helices. *Biochemistry* **34**, 8771-8779.

Villafranca, J.E., Howell, E.E., Oatley, S.J., Xuong, N.-h. i Kraut, J. (1987). An engineered disulfide bond in dihydrofolate reductase. *Biochemistry* **26**, 2182-2189.

Visuri, K. i Nummi, M. (1972). Purification and characterization of crystalline β-amylase from barley. *Eur. J. Biochem.* **28**, 555-565.

Vriend G. (1990). WHAT IF: A molecular modeling and drug design program. *J. Mol. Graph.* **8**, 52-56.

Vtyurin, N. i Panov, V. (1995). Packing constraints of hydrophobic side chains in $(\alpha/\beta)_8$ barrels. *Proteins* **21**, 256-260.

Wang, S., Lue, W. i Chen, J. (1993). Sequence of a β-amylase cDNA from *Zea mays*. Publicat directament a EMBL/GenBank/DDBJ.

ISBN:978-84-694-2163-5/DL:T.885-2011

Warshel, A., Sussman, F. i Hwang, J.K. (1988). Evaluation of catalytic free energies in genetically modified proteins. *J. Mol. Biol.* **201**, 139-159.

Wetzel, R. (1987). Harnessing disulfide bonds using protein engineering. *Trends Biochem. Sci.* **12**, 478-482.

Wilbur, W.J. i Lipman, D.J. (1983). Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. U.S.A.* **80**, 726-730.

Wilmanns, M., Hyde, C.C., Davies, D.R., Kirschner, K. i Jansonius, J.N. (1991). Structural conservation in parallel β/α-barrel enzymes that catalyze three sequential reactions in the pathway of tryptophan biosynthesis. *Biochemistry* **30**, 9161-9169.

Wodak, S.J., Lasters, I., Pio, F. i Claessens, M. (1990). Basic design features of the parallel αβ barrel, a ubiquitous protein-folding motif. *Biochem. Soc. Symp.* **57**, 99-121.

Wodak, S.J., Pontius, J., Vaguine, A. i Richelle, J. (1995). Validating protein structures: from consistency checking to quality assessment in making the most of your model. A: "Proceedings of the CCP4 Study Weekend 6-7 January 1995". Daresbury (UK): Chadwick Library (Hunter, W.N., Thornton, J.M. i Bailey, S., eds), 41-51.

Yang, Z. (1995). Evaluation of several methods for estimating plylogenetic trees when substitution rates differ over nucleotide sites. *J. Mol. Evol.* **40**, 689-697.

Yoshida, N. i Nakamura, K. (1991). Molecular cloning and expression in *Escherichia coli* of cDNA encoding the subunit of sweet potato β-amylase. *J. Biochem.* **110**, 196-201.

Yoshida, N., Hayashi, K. i Nakamura, K. (1992). A nuclear gene encoding β-amylase of sweet potato. *Gene* **120**, 255-259.

Yoshigi, N., Okada, Y., Sahara, H. i Koshino, S. (1994). Expression in *Escherichia coli* of cDNA encoding barley β-amylase and properties of recombinant β-amylase. *Biosci. Biotechnol. Biochem.* **58**, 1080-1086.

Yoshigi, N., Okada, Y., Sahara, H. i Koshino, S. (1994). PCR cloning and sequencing of the β-amylase cDNA from barley. *J. Biochem.* **115**, 47-51.

ISBN:978-84-694-2163-5/DL:T.885-2011

Yutani, K., Ogasahara, K. i Sugino, Y. (1980). pH dependence of stability of the wild-type tryptophan synthase α-subunit and two mutant proteins (Glu_{49} replaced by Met or Gln). *J. Mol. Biol.* **144**, 455-465.

Ziegler, P., Daussant, J. i Loos, K. (1994). Development of β-amylase activity and polymorphism in wheat seedling shoot tissues. *J. Exp. Bot.* **45**, 1147-1155.